7-2012

A Study of Headway Maintenance for Bus Routes: Causes and Effects of "Bus Bunching" in Extensive and Congested Service Areas

Miguel A. Figliozzi
Portland State University, figliozzi@pdx.edu

Wu-chi Feng
Portland State University

Gerardo Lafferriere
Portland State University, Gerardol@pdx.edu

Wei Feng
Portland State University

Citation Details

Figliozzi, Miguel A., Wu-chi Feng, Gerardo Lafferriere, and Wei Feng. A Study of Headway Maintenance for Bus Routes: Causes and Effects of "Bus Bunching" in Extensive and Congested Service Areas. OTREC-RR-12-09.Portland, OR: Transportation Research and Education Center (TREC), 2012. http://dx.doi.org/10.15760/trec.107

OTREC

FINAL REPORT

OREGON
TRANSPORTATION
RESEARCH AND
EDUCATION CONSORTIUM

A Study of Headway Maintenance for Bus Routes: Causes and Effects of "Bus Bunching" in Extensive and Congested Service Areas

OTREC-RR-12-09
July 2012

A National University Transportation Center sponsored by the U.S. Department of Transportation's Research and Innovative Technology Administration

A STUDY OF HEADWAY MAINTENANCE FOR BUS ROUTES: CAUSES AND EFFECTS OF "BUS BUNCHING" IN EXTENSIVE AND CONGESTED SERVICE AREAS

Final Report

OTREC-RR-12-09

by

Miguel Figliozzi, Ph.D.
Wu-chi Feng, Ph.D.
Gerardo Laferriere, Ph.D.
Wei Feng

Portland State University

for

OREGON TRANSPORTATION RESEARCH AND EDUCATION CONSORTIUM

P.O. Box 751
Portland, OR 97207

July 2012

Technical Report Documentation Page

1. Report No. OTREC-RR-12-09	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle A Study of Headway Maintenance for Bus Routes: Causes and Effects of "Bus Bunching" in Extensive and Congested Service Areas		5. Report Date July 2012 6. Performing Organization Code
7. Author(s) Miguel Figliozzi Wu-chi Feng Gerardo Laferriere		8. Performing Organization Report No.
9. Performing Organization Name and Address Portland State University Civil and Environmental Engineering P.O. Box 751 Portland, OR, 97207		10. Work Unit No. (TRAIS) 11. Contract or Grant No. 2010-370
12. Sponsoring Agency Name and Address Oregon Transportation Research and Education Consortium (OTREC) P.O. Box 751 Portland, Oregon 97207		13. Type of Report and Period Covered 14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

A healthy and efficient public transit system is indispensable to reduce congestion, emissions, energy consumption, and car dependency in urban areas. The objective of this research is to 1) develop methods to evaluate and visualize bus service reliability for transit agencies in various temporal and spatial aggregation levels; 2) identify the recurrent unreliability trends of bus routes (focusing on high-frequency service periods) and understand their characteristics, causes and effects; and 3) model service times using linear regression models.
This research utilized six months of archived automatic vehicle location (AVL) and automatic passenger count (APC) data from a low-performance route (Route 15) of TriMet, the public transit provider in the Portland metropolitan area. Route 15 has experienced difficulties in terms of schedule adherence and headway regularity. This research developed methods to summarize causes of bus bunching. We first determined the frequency of each cause (expressed as percentages) meeting pre-determined thresholds. Next, we performed a sensitivity analysis to demonstrate how cause percentage results change using varying difficulty levels of bus bunching thresholds. Finally, we investigated how cause percentage results vary spatially along different route segments. This research also developed novel ways to summarize and visualize vast amounts of bus route operations data in an insightful and intuitive manner: 1) a route/stop level visualization performance measure framework using color contour diagrams and 2) a dynamic interactive bus monitoring visualization framework based on a Google Maps platform. Visualizations proposed in this study can aid transit agency managers and operators to identify operational problems and better understand how such problems propagate spatially and temporally across routes. Finally, regression models were estimated to understand the key factors impacting dwell and travel times.

17. Key Words Visualization, bus bunching, reliability, regression analysis, performance measures	18. Distribution Statement No restrictions. Copies available from OTREC: www.otrec.us

19. Security Classification (of this report)	20. Security Classification (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	68	

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the Oregon Transportation Research and Education Consortium (OTREC) for sponsoring this project. A special acknowledgment for Wei Feng and Scott Price, graduate students from the Civil Engineering and Computer Science departments, respectively, at Portland State University. This project would not have been possible without their hard work, dedication and contributions to transit performance measures, and developing the dynamic visualization application of bus operations performance. The authors are thankful to Steve Callas and David Crout from TriMet for providing data and helpful comments. The authors also are grateful to the reviewers and comments from Intelligent Transportation Systems (ITS) lab students Kristina Hostetler, Eric Albright, Jesse Boudart and many others. Any errors or omissions are the sole responsibility of the authors.

DISCLAIMER

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

A healthy and efficient public transit system is indispensable to reduce congestion, emissions, energy consumption, and car dependency in urban areas. However, low service quality makes transit unattractive for users of private transportation, especially in areas served by bus routes that chronically underperform. In a stochastic environment, deviations from schedules are unavoidable. Uncertain travel times and passenger demand preclude schedule adherence and headway uniformity. A late bus usually encounters more passengers and the extra passengers may create further delays. Meanwhile, if the following bus encounters fewer passengers it tends to run faster. If two buses become too close, "bus bunching" takes place. Bus bunching is associated with longer waiting times for some riders, uneven passenger distribution, overcrowding in late buses, and an overall decrease of level of service and capacity.

The objective of this research is to 1) develop methods to evaluate and visualize bus service reliability for transit agencies in various temporal and spatial aggregation levels; 2) identify the recurrent unreliability trends of bus routes (focusing on high-frequency service periods) and understand their characteristics, causes and effects; and 3) model service times using linear regression models.

This research utilized six months of archived automatic vehicle location (AVL) and automatic passenger count (APC) data from a low-performance route (Route 15) of TriMet, the public transit provider in the Portland metropolitan area. Route 15 has experienced difficulties in terms of schedule adherence and headway regularity. It is characterized by long travel distances across congested areas and high buffer times to absorb delays and variability. Route 15 runs east-west, crossing downtown Portland, with the east terminal located at the Gateway Transit Center and two terminals located at the west end of the route: 1) Montgomery Park and 2) NW Thurman & 27^{th}.

This research developed methods to summarize causes of bus bunching. We first determined the frequency of each cause (expressed as percentages) meeting pre-determined thresholds. Next, we performed a sensitivity analysis to demonstrate how cause percentage results change using varying difficulty levels of bus bunching thresholds. Finally, we investigated how cause percentage results vary spatially along different route segments. This research also developed novel ways to summarize and visualize vast amounts of bus route operations data in an insightful and intuitive manner: 1) a route/stop level visualization performance measure framework using color contour diagrams and 2) a dynamic interactive bus monitoring visualization framework based on a Google Maps platform. Visualizations proposed in this study can aid transit agency managers and operators to identify operational problems and better understand how such problems propagate spatially and temporally across routes.

Finally, regression models were estimated to understand the key factors impacting dwell and travel times.

1.0 INTRODUCTION

1.1 BACKGROUND

A healthy and efficient public transit system is indispensable to reduce congestion, emissions, energy consumption, and car dependency in urban areas. However, low service quality makes transit unattractive to users compared to private transportation. This is true in areas served by bus routes that chronically underperform.

In a stochastic environment, deviations from schedules are unavoidable. Uncertain travel times and passenger demand preclude schedule adherence and headway uniformity. A late bus usually encounters more passengers and the extra passengers may create further delays. Meanwhile, if the following bus encounters fewer passengers it tends to run faster. If two buses become too close, "bus bunching" takes place. Bus bunching is associated with longer waiting times for some riders, uneven passenger distribution, overcrowding in the leading bus, and an overall decrease of level of service and capacity. Therefore, it is important for transit agencies to identify and understand the mechanics of bus bunching, especially recurrent bus bunching during certain time periods or route segments, as it may be due to a scheduling issue rather than an operational problem.

With the implementation of archived automatic vehicle location (AVL) and automatic passenger count (APC) data technology in many transit agencies, bus operational data can be used to help transit agencies monitor performance measures, evaluate service standards, identify potential problems, understand problem characteristics, propose strategies to improve service reliability, and develop visualization tools to better present large amounts of data.

1.2 OBJECTIVES

The objective of this research is to 1) develop methods to evaluate and visualize bus service reliability for transit agencies in various temporal and spatial aggregation levels; 2) identify the recurrent unreliability trends of bus routes (focusing on high-frequency service periods) and understand their characteristics, causes and effects; and 3) model service times using linear regression models.

1.3 REPORT ORGANIZATION

This report is organized as follows: A literature review is provided in Section 2. Section 3 describes the route configuration and archived bus operations data. Section 4 introduces the interactive visualization framework. Section 5 presents bus bunching distribution and causes/effects analysis. Section 6 estimates three regression models for dwell time, travel time and headway. Finally, concluding thoughts and future implications are provided in Section 7.

2.0 LITERATURE REVIEW

Buses are expected to run according to some pre-determined schedules. However, in a stochastic environment, uncertain travel times and passenger demand preclude schedule adherence and headway uniformity. Unreliable transit service not only affects passengers' level of service (increased waiting time and in-vehicle travel time, overcrowding, etc.), but also reduces efficiency and productivity of transit agencies.

2.1 EVALUATION OF TRANSIT SERVICE RELIABILITY

The ability to accurately and effectively analyze various performance measures is fundamental for a transit agency to determine how well it is adhering to its service standards. Transit agencies have a keen interest in understanding methods of generating and displaying transit operations performance data that contain detailed and accurate information.

A comprehensive set of transit reliability indicators can be found in Abkiwitz, Waksam, Englisher & Wilson (1978). Until the implementation of AVL and APC systems to many large transit agencies, substantial detailed operational data are available to compute various performance levels (Bertini & El-Geneidy, 2003; Furth, Hemily, Muller & Strathman, 2006; TCQSM 2nd Eds, 2003). The most widely used reliability indicators in practice are "on-time performance" for low-frequency service (usually headways longer than 10 minutes) and "coefficients of variation of headway deviation" for high-frequency service (TCQSM 2nd Eds, 2003). These indicators are highly aggregated, lack detailed information and therefore are limited in explaining causes of unreliable service. For example, "on-time performance" cannot tell separate proportions of early and late departures, and "coefficients of variation of headway deviation" cannot tell when percentages of extreme headways (bus bunching or large gap) exceed some thresholds. To identify particular unreliable service problems, such as bus bunching temporal and spatial distributions in high-frequency service, and understand their characteristics, causes and effects, more specific techniques and methods are needed in addition to merely measuring reliability performance indicators (e.g., bunching occurrence counts for each stop along a route for each hour of a day, using departure headway as a threshold to define bus bunching (Feng & Figliozzi, 2011)).

2.2 CAUSES OF UNRELIABLE TRANSIT SERVICE

In general, any unreliable transit service is caused by variability and uncertainty, either between-stops or at-stop uncertainty/variability. The sources of stochasticity and variability include passenger-demand uncertainty, driver-behavior uncertainty, traffic congestion, traffic accidents or incidents, and delays at traffic signals. A comprehensive summary of possible causes of unreliable transit service can be found in Abkiwitz, Waksam, Englisher & Wilson (1978); Turnquist (1981); Levinson (1991); and Ceder (2007). On the other hand, in practice schedulers usually set scheduled departure times at the terminals and time points along the route, and insert slack time into these schedules to absorb uncertainty and enhance reliability. However, too much slack time may increase the probability of early arrival and on-time departure (with holding), but can also reduce operating speed and increase operating costs and on-board passenger waiting times at stops (Furth & Muller, 2007).

With the availability of massive AVL-APC data, quantitative analyses of unreliability can be performed by estimating trip time, segment travel time, schedule and headway delay, and their variances or coefficients of variances as functions of a number of explainable variables through regression techniques. Some useful references include: (Bertini & El-Geneidy, 2004; Strathman J. , Dueker, Kimpel, Gerhart, Turner & Taylor, 1999; Strathman J. Dueker, Kimpel, Gerhart, Turner & Taylor, *Service Reliability Impacts of Computer-aided Dispatching and Automatic Vehicle Location Technology: A Tri-Met Case Study*, 2000; Strathman, Kimpel & Callas, 2003; Dueker, Kimpel & Strathman, 2004; Kimpel, Strathman & Callas, 2008; Kimpel, 2001; El-Geneidy, Horning & Krizek, 2010). Research results using linear regression models have showed that significant factors are passenger-demand variability and bus operators' driving behavior. Although regression models can explain the significance between various explanatory variables and certain dependent variables of interest, they are not able to explain the formation and dissipation of bus bunching. Also, explanatory variables such as holding time at time points, number of signalized intersections, and transit signal priority requests are not considered in the above mentioned studies.

Two studies in the literature have focused on bus bunching causes. Hammerle, Haynes & McNeil (2005) analyzed two causes of bus bunching: on-route effects and at the terminal or route start. By analyzing a two-day sample of archived operations data, they found that most bus bunching was the result of irregular departure headways at the terminals instead of on-route effects. Time-space trajectory graphs of several bus trips were plotted to help identify bus bunching problems and causes. Feng & Figliozzi (2011) analyzed the relationship between bus bunching and stop-level causes. They found that short departure headway at a stop is mainly due to irregular departure from an upstream stop instead of irregular passenger demand or uncertain travel times between each two consecutive stops.

2.3 VISUALIZATION OF BUS OPERATIONS DATA

A number of low-cost surveillance, monitoring and management systems as part of intelligent transportation systems (ITS) programs now exist, enabling transit agencies to collect advanced operational data for testing and analyzing operating efficiency and service reliability (Furth P. , 2000; Furth, Hemily, Muller & Strathman, 2006; Bertini & El-Geneidy, 2003; Berkow, Chee, Bertini & Monsere, 2007). Although the availability of such rich archived AVL/APC data makes it possible for transit agencies to generate valuable transit performance measures, a large amount of useful information is underutilized due to the sheer volume of information available. This vast amount of data creates the need to use visualization techniques that can easily convey key performance measures. Kimpel (2006) addresses the potential of overall data visualization for enhancing exploratory analysis, pattern identification and hypothesis development. Other transit data visualization examples offered by Kimpel include general mapping of quantity information, linear referencing, time-distance diagrams and 3-D visualization. Berkow, EI-Geneidy, Bertini & Crout (2009) also investigate the power of data visualization to understand the capacity for BDS data. Their research focuses on using new aspects of performance measurement and visualization to conduct hierarchical analysis over a one-year period, starting with system-level analysis and ending with stop-level analysis. They employed GIS and statistical modeling to demonstrate performance measures. However, most of the visualization tools are examples for illustration. None of them start from the prospect of an interactive interface for users (transit agencies) to report performance measures according to user inputs so that results can be directly generated from an archived database through embedded data processing techniques and algorithms. Liao & Liu (2010) developed a data processing framework that provides users with an interactive interface to select any time point-level or route-level performance measures that can be directly read or computed from the dataset (one-month sample).

3.0 ROUTE CONFIGURATION AND DATA DESCRIPTION

3.1 ROUTE CONFIGURATION

This study focuses on TriMet's Route 15, which experiences difficulties in terms of schedule adherence and headway regularity. Route 15 is characterized by long travel distances across congested areas and high buffer times to absorb delays and variability. Route 15 runs east-west, crossing downtown Portland, with the east terminal located at the Gateway Transit Center and two terminals located at the west end of the route: 1) Montgomery Park and 2) NW Thurman & 27th. Figure 3-1 and Table 3-1 show the route schematic as well as the key stop names for both westbound and eastbound services.

Time points (stops with scheduled departure times) are depicted by the numbered circles in the route schematic of Figure 3-1; white circles with black numbers indicate stops for the westbound route, and black circles with white numbers indicate stops along the eastbound route. Table 3-1 lists names of all time points along the route. For the majority of the day, Route 15 is in low-frequency service where headways between buses are approximately 15 minutes. However, in the a.m. peak hours, westbound passenger demand is much higher than other times of day due to morning commute to work in downtown Portland. Therefore, additional short trips are added to the route in the a.m. peak hours for westbound travel. These additional short trips run from the stop at SE Stark & 93rd to the stop at SW Morrison & 17th, and therefore reduce the departure headways of stops within this zone to five to eight minutes, which is so-called high-frequency service. Similar additional short trips are added to the eastbound travel direction during the p.m. peak hours due to evening commute to home from downtown Portland. In eastbound high-frequency service, additional short trips start from SW Salmon & 5th, but may end at any of the three downstream time points, SE Belmont & 39th, SE Belmont & 60th, or SE Washington & 82nd. The approximate time periods for high-frequency service are shown in Figure 3-1, 6-10 a.m. for westbound and 4-7 p.m. for eastbound. The time points that are within the high-frequency zone of the route are indicated in bold typeface in Table 3-1.

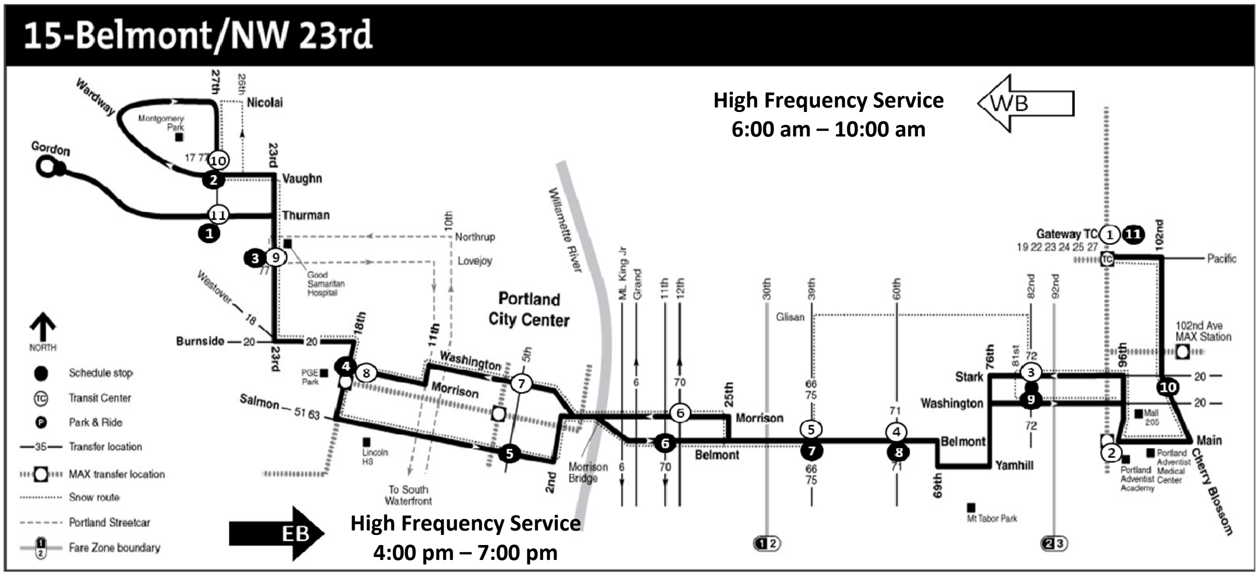


Figure 3-1. Route 15 schematic (source TriMet website)

Table 3-1. Route 15 Time Points

Eastbound	***Westbound***
1. NW Thurman & 27th	1. Gateway TC
2. Montgomery Park	2. SE 102nd & Washington
3. NW 23rd & Marshall	**3. SE Stark & 82nd**
4. SW 18th & Morrison	**4. SE Belmont & 60th**
5. SW Salmon & 5th	**5. SE Belmont & 39th**
6. SE Belmont & 11th	**6. SE Morrison & 12th**
7. SE Belmont & 39th	**7. SW Washington & 5th**
8. SE Belmont & 60th	**8. SW Morrison & 17th**
9. SE Washington & 82nd	9. NW 23rd & Lovejoy
10. SE Washington & 103rd	10. Montgomery Park
11. Gateway TC	11. NW Thurman & 27th

3.2 DATA DESCRIPTION

A number of low-cost surveillance, monitoring and management systems as part of ITS programs now exist, enabling transit agencies to collect advanced operational data for testing and analysis of operating efficiency and service reliability (Furth, Hemily, Muller & Strathman, 2006; Furth P., 2000). TriMet has a long history of employing these technologies, particularly the use of the Bus Dispatching Systems (BDS). Strathman et al. (2001) described the usage of automated BDS data based on GPS-based AVL-APC technology, dead reckoning sensors, and voice and data communication within a mobile radio system. In addition to the aforementioned technologies, each TriMet bus has an on-board computer and a control head displaying schedule adherence to drivers, detection and reporting of schedule, and route deviations to dispatchers, and two-way, pre-programmed messaging between drivers and dispatchers. The BDS implemented by TriMet collects and archives stop-level data as part of its overall service control and management system. Regularly scheduled weekday service contains approximately 10,000 trips, and on a typical weekday the BDS records over 500,000 entries (Berkow, EI-Geneidy, Bertini & Crout, 2009). A sample of the archived stop-event data for Route 15 is shown in Table 3-2.

Table 3-2. Stop-Event Data Sample of Route 15

Date	Leave _time	Train	Stop _time	Arrive _time	Dwell	Stop_id	Door	Lift	ons	offs	Load	Mileage
9/14/2009	21150	1501	21120	21136	0	8989	0	0	0	0	2	8.1
9/14/2009	21216	1501	21194	21182	10	7162	1	0	2	0	4	8.3
9/14/2009	21262	1501	21248	21238	7	8963	1	0	2	1	5	8.5
9/14/2009	21294	1501	21286	21278	0	7174	0	0	0	0	5	8.6
9/14/2009	21344	1501	21327	21320	6	718	2	0	1	0	6	8.7
9/14/2009	21384	1501	21373	21360	0	749	0	0	0	0	6	8.8
9/14/2009	21430	1501	21407	21394	5	8511	1	0	1	0	8	8.9
9/14/2009	21496	1501	21480	21472	8	6911	2	0	0	1	7	9.1
9/14/2009	21590	1501	21575	21582	0	5016	0	0	0	0	7	9.3
9/14/2009	21636	1501	21611	21602	0	5014	0	0	0	0	7	9.4
…	…	…	…	…	…	…	…	…	…	…	…	…

Whenever a bus arrives at or departs from a stop, a new record is entered. The column “leave_time” is the actual departure time of that bus from that stop; “stop_time” is the scheduled departure time of that bus at that stop; and “arrive_time” is the actual arrival time for that bus at that stop, all of which is expressed in seconds after midnight. There is no scheduled “arrive time” for any stop. Also, the “stop_time” for time points along the route is the real scheduled departure time, for all other stops, this “stop_time” is interpolated. The dwell time here is recorded as the time (in seconds) that the door is open; therefore, dwell time is usually smaller than the actual departure time minus actual arrival time. All the other information shown in Table 3-2 is self-explanatory. Note that there are additional columns in the stop-event data which are not shown in this sample. These include route number, direction, x-y coordinates, etc.

The studied data period ranges from Sept. 14, 2009, to Feb. 26, 2010, including all weekdays (totaling 115 days).

4.0 EVALUATION AND VISUALIZATION OF BUS OPERATIONS AND SERVICE RELIABILITY

This section demonstrates methodologies that: 1) convert bus operational records of any specific date and time period into visible outputs; and 2) generate and visualize reliability performance measures aggregated over a number of days.

The first part (section 4.1) focuses on visualizing the most detailed operational history of all buses on a selected route, and therefore it is usually appropriate to show only a specific time period within a specific date, according to a user's input. Two visualization tools are developed in this section: static visualization (time-space diagram) and dynamic visualization (a Google Maps-based application).

The second part (section 4.2) focuses on evaluation and visualization of bus service reliability. Performance measures are usually some statistics calculated from a large sample of bus operations data and sample size can be determined by users, which may contain data of several days or months. Two typical performance measures for low- and high-frequency service suggested by (TCQSM 2nd Eds, 2003) are shown in a two-dimensional (temporal-spatial) diagram.

4.1 BUS OPERATIONS VISUALIZATION

4.1.1 Static Visualization

Time-space diagrams are a very useful tool to visualize large amounts of tabular data. They are very helpful to identify bus operations and scheduling problems and evaluate the effectiveness of management interventions (Liao & Liu, 2010; Hranac, Kwon, Bachmann & Petty, 2011). However, this tool is mainly used by researchers to analyze particular problems in a specific time of day, as an example. It will be very convenient if a data processing framework is developed in which users (transit agencies) can select any route, travel direction, date and time period (as long as it is within the archived database), and simply click a button to cause a time-space diagram to pop up showing the scheduled and actual trajectories of all buses running along that route and direction during that time period.

An example of a time-space diagram is shown in Figure 4-1. To generate this diagram, users only have to select the date Sept. 16, 2009, westbound travel direction in the a.m. peak hours between 6:30 and 9:30 a.m., and click a button. The X-axis represents time and the Y-axis shows time point names and distances to terminal. The solid lines represent actual travel trajectories and dashed lines represent scheduled travel trajectories; trips are separated by colors. Trip numbers and relative colors are shown on the right in the legend. It is very easy to identify some bus bunching trips from this diagram.

Although the time-space diagram can show movements of all buses along a route in one direction in a certain time of day, detailed information about each bus are not available. For example, how have two buses become bunched? What are their schedule adherence and departure headways at each stop, passenger boarding and alighting activities, and number of on-board passengers?

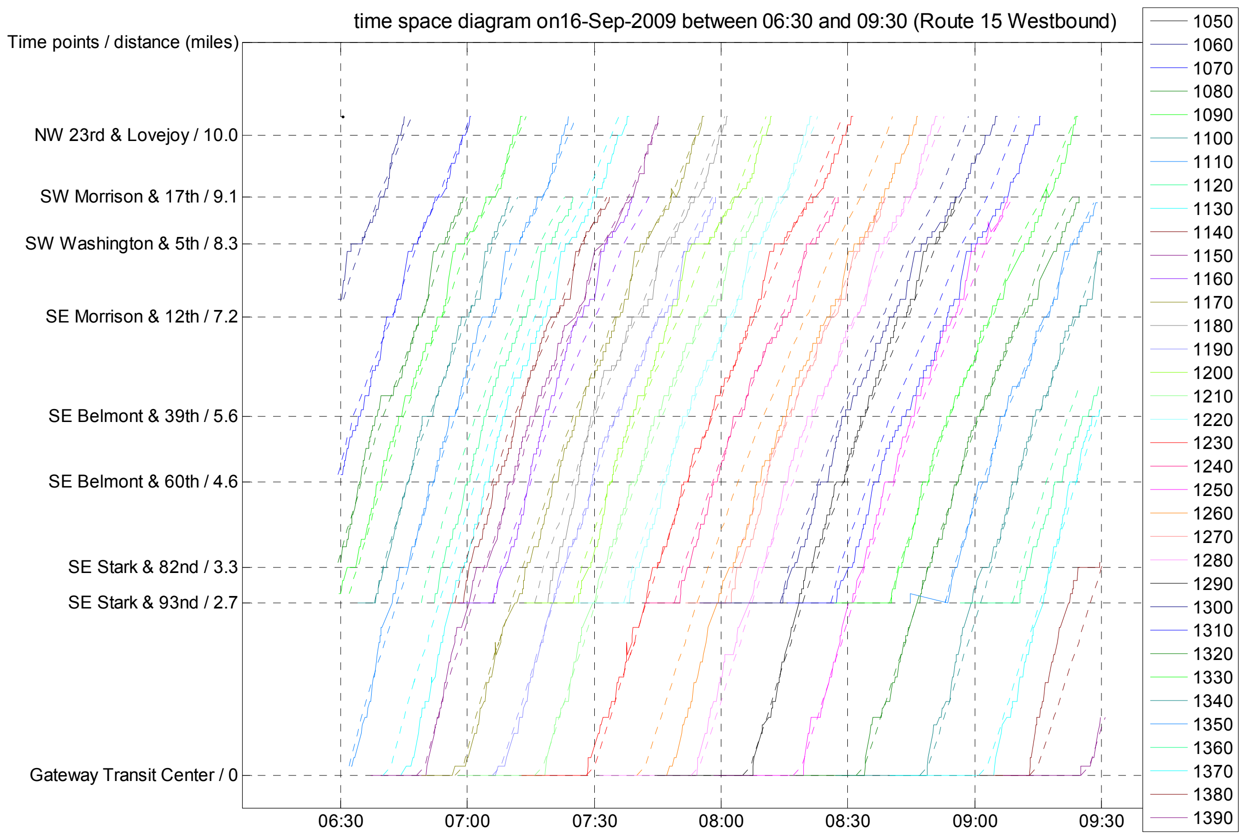


Figure 4-1. Time-space diagram example

4.1.2 Dynamic Visualization

One of the issues with static analysis of historical data is the difficulty in understanding some of the finer details of what is happening within the system. That is, it tends to provide a macro-level view of the system. Therefore, we argue that both macro-level as well as more detailed micro-level information are necessary components to understand such complex systems. However, the overwhelming amount of micro-level data for such large-scale systems can lead to looking at irrelevant data. For our TriMet Route 15 study, we focus on enabling time-varying display of the archived data.

This dynamic visualization system consists of three primary components: an SQL database, PHP-based web server, and the web-based JavaScript client. We use an SQL database to store the BDS data. The data is accessed through the web interface. Initially, our implementation had much of the functionality using only the web server; yet we found that the interactions, particularly updates to the visualization, were not scalable and moved towards a client-based approach. Thus, the web server in our implementation responds to single-day queries issued by a web client. The server then translates the request into an SQL query and provides the data inline to the client. The client web browser directly handles much of the visualization interaction without having to reconnect with the server. Using this approach allows fairly scalable implementation with much of the work being handled at the client.

At the client, JavaScript is used to parse additional information and calculations, and to store the main data structures. Once the information is loaded into JavaScript data structures, the Google API is called to perform the actual mapping on to the map. We used customized icons (in this case, buses) we designed and loaded them on the Google map as clickable events. The user can then run the visualization of the day's event, looking at a bus's particular location during the day. Clicking on a single bus (see Figure 4-2) on the map brings up additional information relevant to the particular bus, such as schedule adherence, vehicle number and headway. The user can "play" the movement of the buses during the day and drill down into a particular bus to see more details. Finally, the icons were designed to provide the user visual information, such as bus capacity and performance, and will be described shortly.

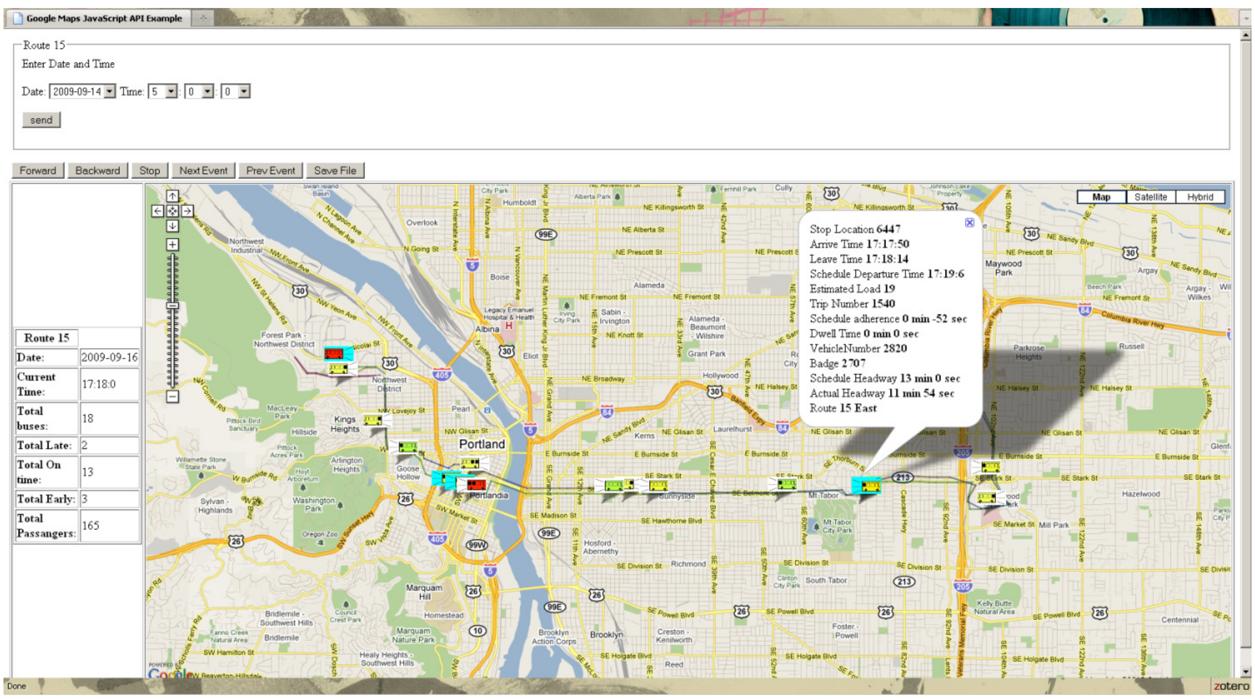


Figure 4-2. Dynamic visualization framework interface snapshot for TriMet Route 15

Figure 4-2 provides a snapshot of the framework showing all buses that are running along the route at this time. On the upper left corner are options for users to select which day and time period to "play." This "play" function provides both "forward" and "backward" moving options for users to observe how bus bunching propagates over time and space. The time period to "play" can either be a fixed time interval (which should be larger than the load time – five seconds in this case) or by an event-based moving interval (i.e., whenever there is a bus arrival or departure activity, reload the map.) Double clicking the "Forward" or "Backward" button can speed up the animation speed. The information displayed in the pop-up window is pulled directly from the database or by using basic math functions. Table 4-1 provides further details of additional map features.

Table 4-1. Descriptions of Map Features

Route Map	Direction Indicator	Estimated Passenger Loads	Bus Stop Service Indicator	On-Time Performance Indicator
• Purple = Westbound • Green = Eastbound	• Distinguished by rear emission • Opposite of emission direction • Also indicated in pop-up window	• Shown using window icons • Number of black windows represent estimated passenger load: 1. ≤ 25% 2. 25% - 50% 3. 50% - 75% 4. >75% • Bus capacity = 60	• Depicted by background color of bus icon • Blue = bus serving stop • White = bus running between stops	• Shown by colors of bus icons • Red = Late; schedule adherence > 5 minutes • Green = Early schedule adherence < -1 minute • Yellow = On time; -1 minute < schedule adherence < 5 minutes

Bus route is depicted on the Google Maps according to TriMet's bus location coordinates. In this pilot example we manually created the routes of both directions by connecting piece-wise linear routes between consecutive stops. One drawback of this is that they may not represent the actual route lines when the route between two consecutive stops is not a straight line. In this map, the purple line represents westbound travel direction and the green line represents eastbound travel direction.

The bus rear is distinguished by the thin lines representing exhaust emissions and wind turbulence at the back of the bus. Running direction is opposite to the emission direction. This information is also shown later in a pop-up window. This information can be read directly from the archived data in the "direction" column. A bus's estimated load is shown by the windows in the bus icons. There are four windows for each bus, from the rear window to the front window; the number (1, 2, 3, 4) of black windows represent the estimated on-board passengers as less than 25%, 25%-50%, 50%-75% and over 75% of the bus capacity (60 in this case), respectively. This information can also be read directly from the database in the "estimate_load" column.

Buses that are serving at a stop are shown by a blue background of the bus icon while a white background means they are running between stops. Since the archived data are stop-event data – there are no data or coordinates for buses running between stops – the bus GPS data are available only when one is at a stop. Therefore, in order to simulate buses running between stops, we used a simple extrapolation method to estimate bus positions. The distance from the previous stop of a running bus will be proportional to the travel time between the previous stop and the next stop. Note that because this is archived historical data, next arrival bus times are known as priori.

Bus on-time performance is shown by color codes. A red color represents that a bus left five minutes later than the scheduled departure time from its previous stop (late departure). Green represents a bus leaving one minute earlier than the schedule departure time from its previous stop (early departure). Lastly, yellow represents a bus departing no earlier than one minute and no later than five minutes than the scheduled departure time from its previous stop (on-time departure). The "previous stop" will be updated every time a bus departs from a stop. So if a bus is running between stops 1 and 2, the "previous stop" is stop 1; if a bus is serving at stop 2, the "previous stop" still means stop 1. Only if the bus departs from stop 2 that the "previous stop" means stop 2. The on-time performance can be simply calculated by "actual leave time – schedule leave time" for each of the current buses. The result of this simple formula is schedule adherence; this performance is also shown in the pop-up window for each bus with values in the unit of seconds.

Each performance measure of every current bus is visually shown in one map. By showing this dynamically, users can have a better overall understanding of how the buses operated temporally and spatially and how they interacted with each other.

For users who want to look at more details about one of the current buses, a pop-up window will be displayed whenever a user clicks the bus icon (also shown in Figure 4-2). The pop-up window will show information including: vehicle number, train, operator badge number, route number, direction, schedule adherence, previous stop_ID, current or next stop_ID, dwell time at a previous stop, schedule and actual headway of the previous stop, and travel time from the previous stop to the next or current stop. The values of vehicle number, train, operator badge number, route number, and direction can be directly read from the database by matching with the selected bus. The methods of calculating schedule adherence and defining the previous stop have been introduced above while the rest are explained below.

The previous stop_ID can be found from the database. The dwell time at the previous stop also can be matched. For current or next stop_ID, first judge if the selected bus is serving at a stop. If yes, then the current stop_ID will be shown; if not, the next stop_ID will be found and shown. Furthermore, once the previous stop_ID and current or next stop_ID have been found, the travel time from previous stop to the current or next stop can be easily calculated from the database by "current or next stop arrive time – previous stop leave time."

The method of calculating headway of the previous stop is as follows: first find the previous stop by the aforementioned method, then the leave time of the selected bus from the previous stop can be found. Second, search for the latest leave time that is smaller than the selected bus leave time and record the bus number, but make sure the bus is different from the selected bus (sometimes there are multiple records for the same bus serving at the same stop.) Finally, the actual headway at the previous stop can be calculated by "leave time of the selected bus – leave time of the last departure bus," and schedule headway at the previous stop can be calculated by "schedule depart time of the selected bus – schedule depart time of the last depart bus." By comparing these two values, users can also have an idea about the headway deviation, and this is especially useful in a high-frequency service period. It can help users to understand how irregular headways are propagating temporally and spatially among buses.

The dynamic visualization is based on archived historical data, and some of the performance measures need records that are later than the current time. It cannot be directly implemented into the real-time bus monitoring system, but most of that information only uses previous time records that are earlier than the current time. Therefore, if new data communication technology can be implemented in TriMet, a real-time bus monitoring system with some real-time operational performance measures can be developed easily based on the work here, and it will help dispatchers to make real-time decisions when there are minor disruptions so that major disruptions can be avoided at the earliest time.

All of the information shown in the pop-up window is updated every time a bus departs from a stop. The actual and schedule headways are the time difference between the current bus (actual and schedule) departure time and the previous bus (actual and schedule) departure time. The ability to dynamically visualize how buses in a route move spatially and temporally is extremely useful for understanding transit operations performance, particularly to better understand how the impacts of decreased on-time performance and increased headway deviations propagate spatially and temporally and result in effects such as bus bunching.

A preliminary version of the dynamic visualization tool is available for use and exploration at the following web address:
http://web.cecs.pdx.edu/~eea/trimetViz/tspviz.php
or
http://web.cecs.pdx.edu/~zaral/portals/trimetportalgraph.php.

The reader should be warned that the systems were developed for Mozilla Firefox browsers version 9.0 and do not work well on Explorer or other browsers.

Implementation Issues

In the implementation of our visualization system, we had several problems. First, the routes of the buses needed to be created in order to have the bus icons follow the appropriate roads on Google Maps. In order to handle this, we manually created the bus routes as piece-wise linear routes. When a bus needs to be displayed between two stops, its icon will follow these routes. Second, the database only records the arrival and departure time of the buses, but does not have actual bus location information between stops. To determine where the bus is in between two stops, we take the time required for the bus to reach the next stop and linearly interpolate its location based on the time and the distance along the piece-wise linear route we created. We hope to incorporate the GPS recorded data into the visualization when they become available.

4.2 BUS SERVICE RELIABILITY EVALUATION AND VISUALIZATION

This section demonstrates a "drill-down systematic" framework for evaluating and visualizing bus service reliability performance utilizing archived bus operational data. It starts with a route-level reliability performance measure that disaggregates into temporal and spatial dimensions, and then provides options to concentrate on any one dimension or a single stop in a particular time period.

The frequency of transit service determines which performance measures ought to be analyzed, as there are particular performance measures which are best applied depending on either low- or high-frequency service (TCQSM 2nd Eds, 2003). For either case, performance measures were computed hourly at first, and for all time points or segments, using archived data for all weekdays over a certain time period according to users' choice (six months data in this example); and they are displayed using color contour diagrams. Examination of the resulting color contour diagrams can reveal whether additional analysis is necessary for a single stop across all hours of a day (temporal analysis); for a single time period across all scheduled stops (spatial analysis); or for a single stop during a single time period. The framework can easily generate distributions and several statistics of various performance measures, such as on-time performance, headway adherence, dwell time, travel time and speed, passenger boarding and alighting, lift use, etc.

4.2.1 Low-Frequency Service

Recall that low-frequency service is determined by headways greater than 10 minutes. Per the (TCQSM 2nd Eds, 2003), on-time performance is one of the most popular reliability measures used in the transit industry and is generally best applied for low-frequency service. While the (TCQSM 2nd Eds, 2003) suggests "on-time" performance as being "0 to 5 minutes late," TriMet defines "on-time" performance as being "no more than 1 minute early to no later than 5 minutes past scheduled departure time." Therefore, the index for on-time performance percentage is calculated using the following formula:

$$1 - \left(\frac{\text{early depart records}}{\text{all depart records}}\right) - \left(\frac{\text{late depart records}}{\text{all depart records}}\right) \quad \textbf{(1)}$$

To apply this formula, we can calculate the hourly on-time performance for each scheduled stop along the route for both directions using the six-month's worth of archived BDS data. The results for the westbound Route 15 are shown in Figure 4-3. The Level of Service (LOS) ranges for on-time performance are as follows:

- A 0.95 - 1.00
- B 0.90 - 0.949
- C 0.85 - 0.899
- D 0.80 - 0.849
- E 0.75 - 0.799
- F < 0.75

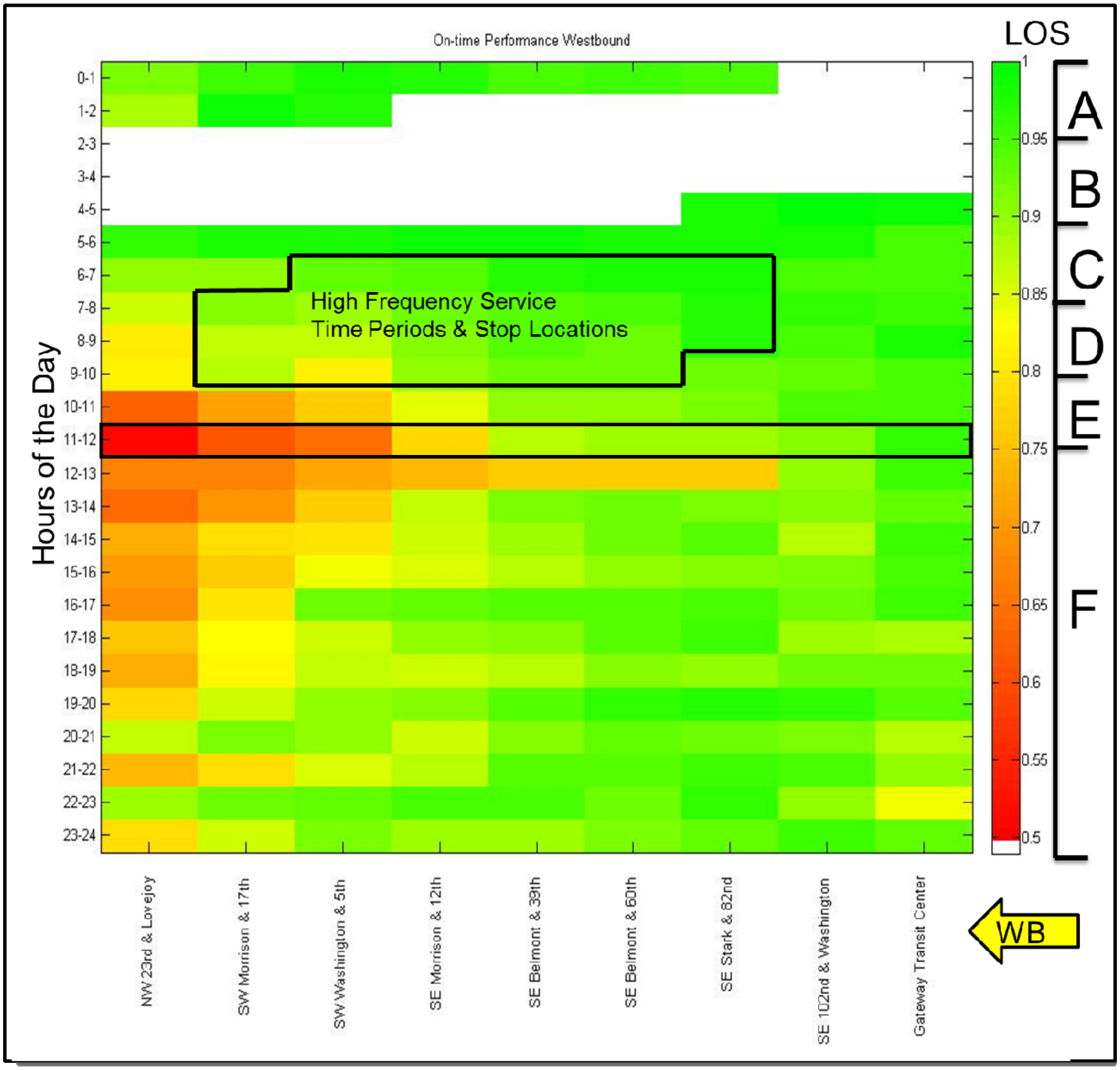


Figure 4-3. On-time performance for Route 15 westbound

In Figure 4-3, on-time performance for nine time points along the westbound (from right to left) Route 15 were calculated hourly. The on-time performance indexes are shown by color, ranging from red to green to represent Low LOS to High LOS. The on-time performance values and their relative colors are shown on the color bar. The white area in the color contour diagram indicates that there is no data for those time periods. The outlined area in the color contour diagram denotes the high-frequency service time periods and stop locations.

When considering spatial factors, the changing colors from green to red horizontally across the diagram in Figure 4-3 suggest that the on-time performance becomes gradually worse from the east terminal to the west end of the route. Temporally, the changing colors vertically indicate the on-time performances are the best at the start of the day from 4-6 a.m.; and they are the worst in the middle of day from 10 a.m. to 1 p.m. There is a 10-hour period (from 10 a.m. to 8 p.m.)

when the stops on the west-side of the river generally have very low on-time performances. Because these west-side stops continue to serve relatively densely populated downtown areas, recurrent problems may exist with poor traffic signaling and transit coordination.

While the color contour diagram in Figure 4-3 shows the overall picture of Route 15's on-time performance for six-month's worth of BDS data broken down for each hour of a day – note that it not always has to be broken down to each hour as it can also be aggregated into a.m. peak, off peak, p.m. peak, and others – it is also worth examining either temporal or spatial points of interest. In this case, the on-time performance for the 11 a.m. to noon time period appears to experience all ranges of LOS, thus requiring additional spatial analysis for this time period. If we were curious about the performance behavior of a single stop for all hours of the day, then temporal analysis would be appropriate. Figure 4-4 illustrates on-time performance data for the 11 a.m. to noon time period using a box-plot graph. Each box-plot shows the range of schedule deviations (actual departure time minus scheduled departure time) for each stop during this time period. Comparing the data range of each box plot to the "on-time" threshold (one minute early to five minutes late, depicted by the two solid black lines running across the graph) illustrates that the proportion of data that lie within the threshold lines. The data range for the stops towards the west-end of the route lie above the "on-time" threshold, indicating poor on-time performance, and particularly more late departures.

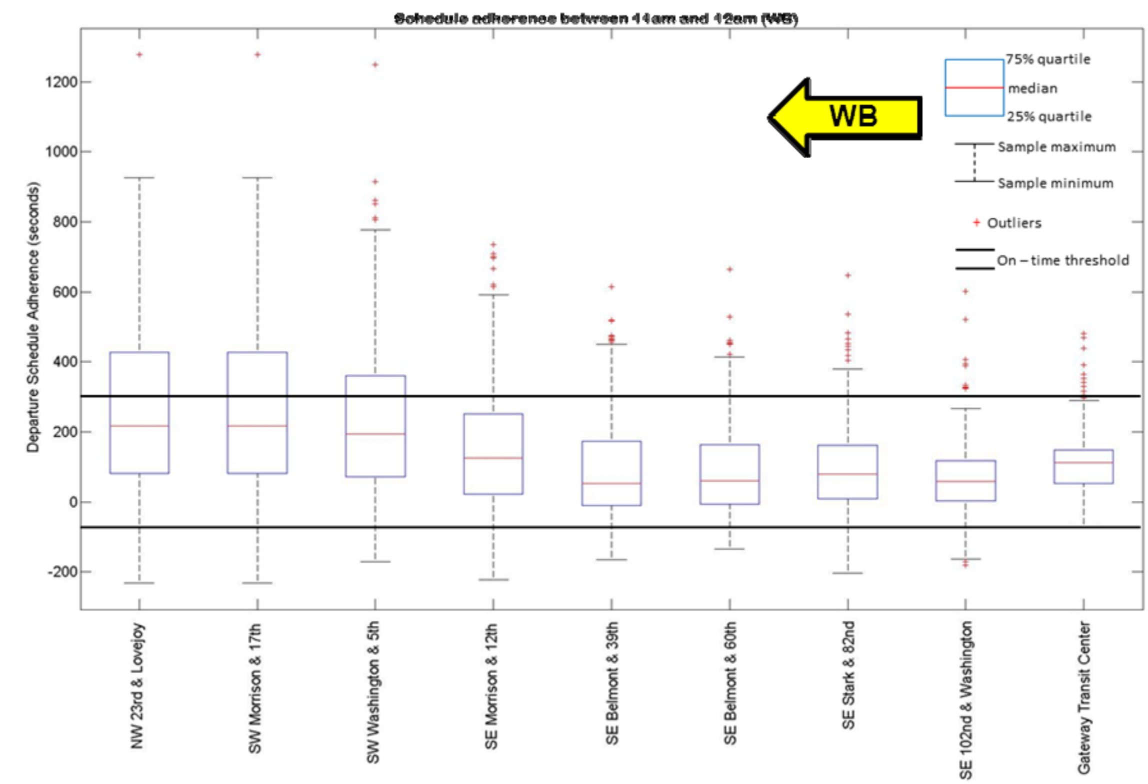


Figure 4-4. Box-plots of schedule deviations for Route 15 westbound

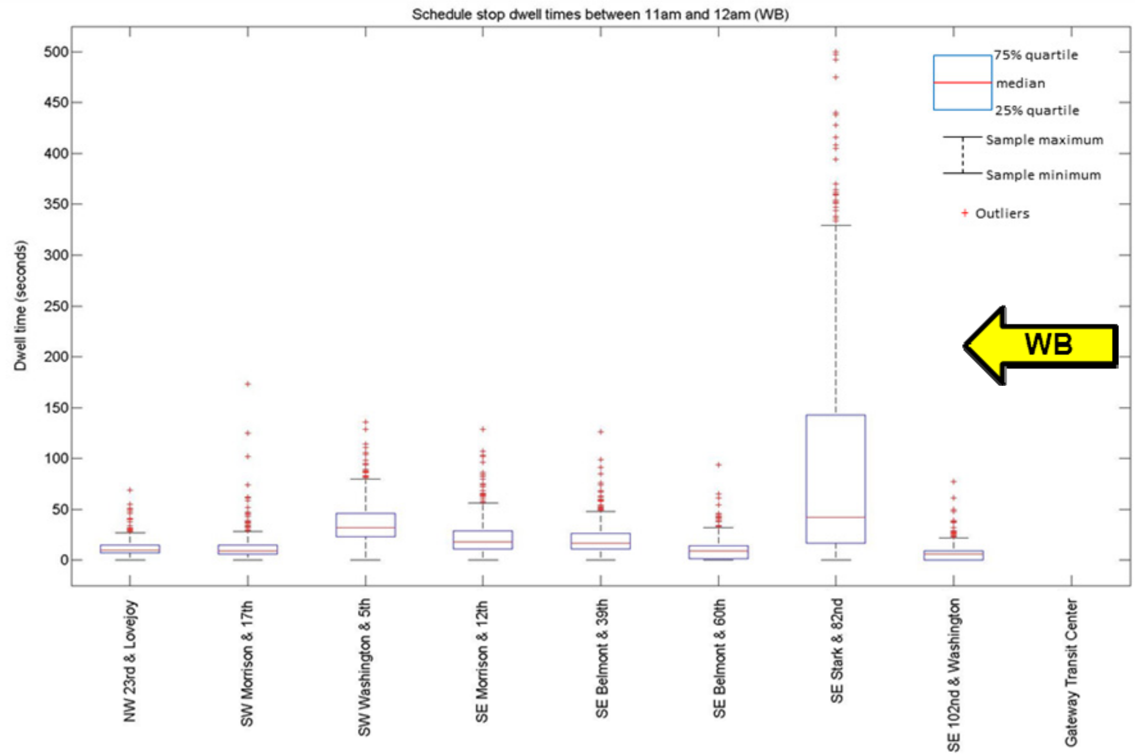


Figure 4-5. Box-plot of dwell time for Route 15 westbound

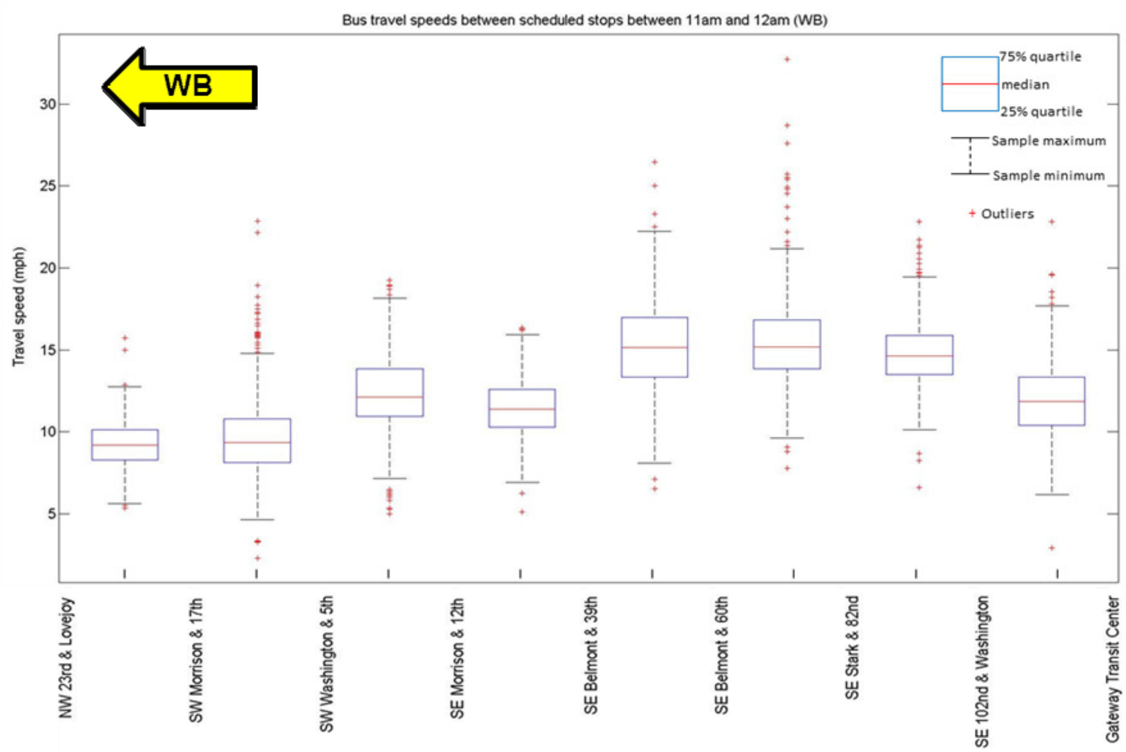


Figure 4-6. Box-plot of travel speed for Route 15 westbound

It is also worth investigating potential causes of poor on-time performance measures. The BDS data readily offers dwell time data for each scheduled stop and segment, and a straightforward way to generate travel time data. In order to provide comparable analysis between each route segment – since each segment is a different length, thus having different travel times – we first converted travel time data to speed in miles per hour. Again using box-plots, we examine dwell times and travel speeds to determine any trends or correlations to on-time performance in Figure 4-5 and Figure 4-6.

Figure 4-5 shows curious behavior, particularly at the SE Stark & 82nd stop. The vertical stretch of this box-plot suggests high variability and longer duration of dwell times at this particular stop during the 11 a.m. to noon time period. It is possible that the bus may arrive early to that particular stop, warranting a longer dwell time in order to get back on schedule (departure time). However, it cannot be readily concluded that the dwell time performance at this stop directly affects the on-time performance; perhaps the events happening at this stop continue to propagate along the route, causing later issues. For all the downstream scheduled stops, dwell time for each stop alone does not appear to be a key factor for the poor on-time performance for that stop. Figure 4-6 shows the speed box-plots for the same time for all the scheduled stops. The median travel speed between SE Belmont & 39th and SE Morrison & 12th, and the travel speed between SE Morrison & 12th and SW Washington & 5th are lower than the upstream median speeds. Also, the median travel speeds for the further downstream segments are even lower, which may possibly cause the large amount of late departure for those downstream stops. Therefore, Figure 4-4, 4-5 and 4-6 indicate that an adjustment of the schedule departure times for the west end time points around this hour.

Finally, zeroing in on a specific problem area and time period is now possible by inspecting the color contour diagram for the low LOS (orange and red) areas outside the high-frequency service area. For example, in Figure 4-3 the stop at SW Washington & 5th between 11 a.m. and noon is depicted by a deep orange color, implying an LOS Grade F for on-time performance. This can be confirmed by plotting the distribution of the schedule adherence records. Figure 4-7 depicts this distribution for the stop at SW Washington & 5th between 11 a.m. and noon.

The dashed line in Figure 4-7 represents the ideal, cumulative distribution line; the curved red line is the actual cumulative distribution line; and the bars denote the frequencies of schedule adherence for each minute, early or late. From this figure we can see that based on six-month's worth of data, from 11 a.m. to noon, the on-time performance measure of this stop lies at 64% (where "on-time" is defined as between one minute early to five minutes late), indeed equating to an LOS Grade F. It is also interesting to note that 95% of the schedule adherences are larger than zero (bars to the right of the ideal line), revealing that most bus departures from this stop (between 11 a.m. to noon) are technically late; and over 30% of them are later than five minutes. Few buses depart early (only 5%), but this may be due to a TriMet operating rule mandating bus drivers to depart according to schedule. If they arrive early, they are to wait until less than one minute earlier than scheduled departure time; however, if they are running late, they are to depart immediately after serving passengers.

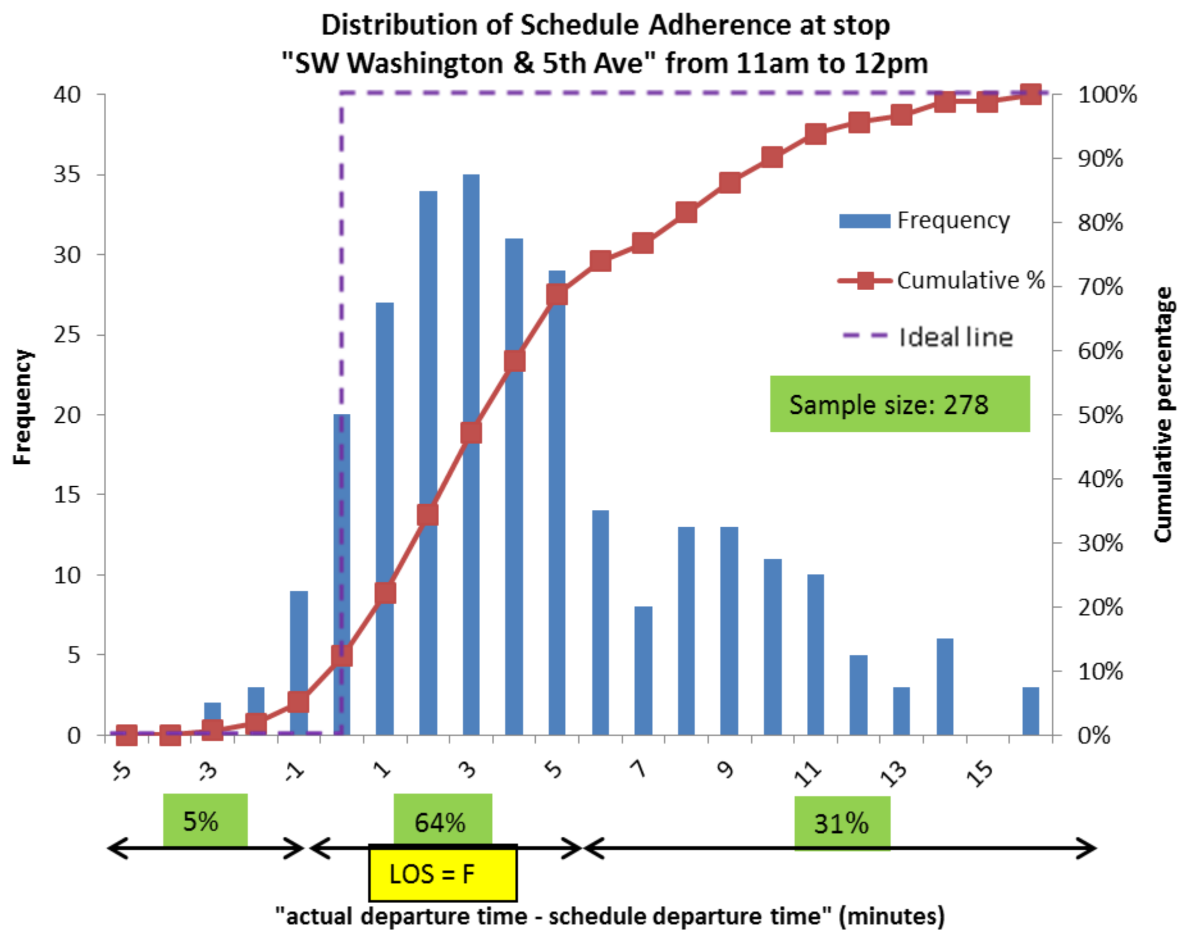


Figure 4-7. Schedule adherence distributions for SW Washington & 5th stop between 11 a.m. to noon

Similar plots and analysis for eastbound Route 15 is not shown for brevity.

4.2.2 High-Frequency Service

As the on-time performance measure is best applied for low-frequency service, headway adherence is used to measure service reliability for high-frequency service routes. Bear in mind, high-frequency service is defined by headways that are generally 10 minutes or less. The formula for calculating headway adherence is shown below (TCQSM 2nd Eds, 2003):

$$c_{vh} = \left(\frac{\text{standard deviation of headway deviations}}{\text{mean scheduled headway}}\right) \qquad \textbf{(2)}$$

Where c_{vh} is the coefficient of variation of headways and headway deviation is the difference between the actual headway and the scheduled headway. The headway adherence values and relative LOS are below:

- A 0.00-0.21
- B 0.22-0.30

- C   0.31-0.39
- D   0.40-0.52
- E   0.53-0.74
- F   > 0.75

By using the six-month's worth of archived BDS data and applying this formula, we calculated hourly headway adherence for each time point along the route for both directions. The results for the westbound Route 15 are shown in Figure 4-8.

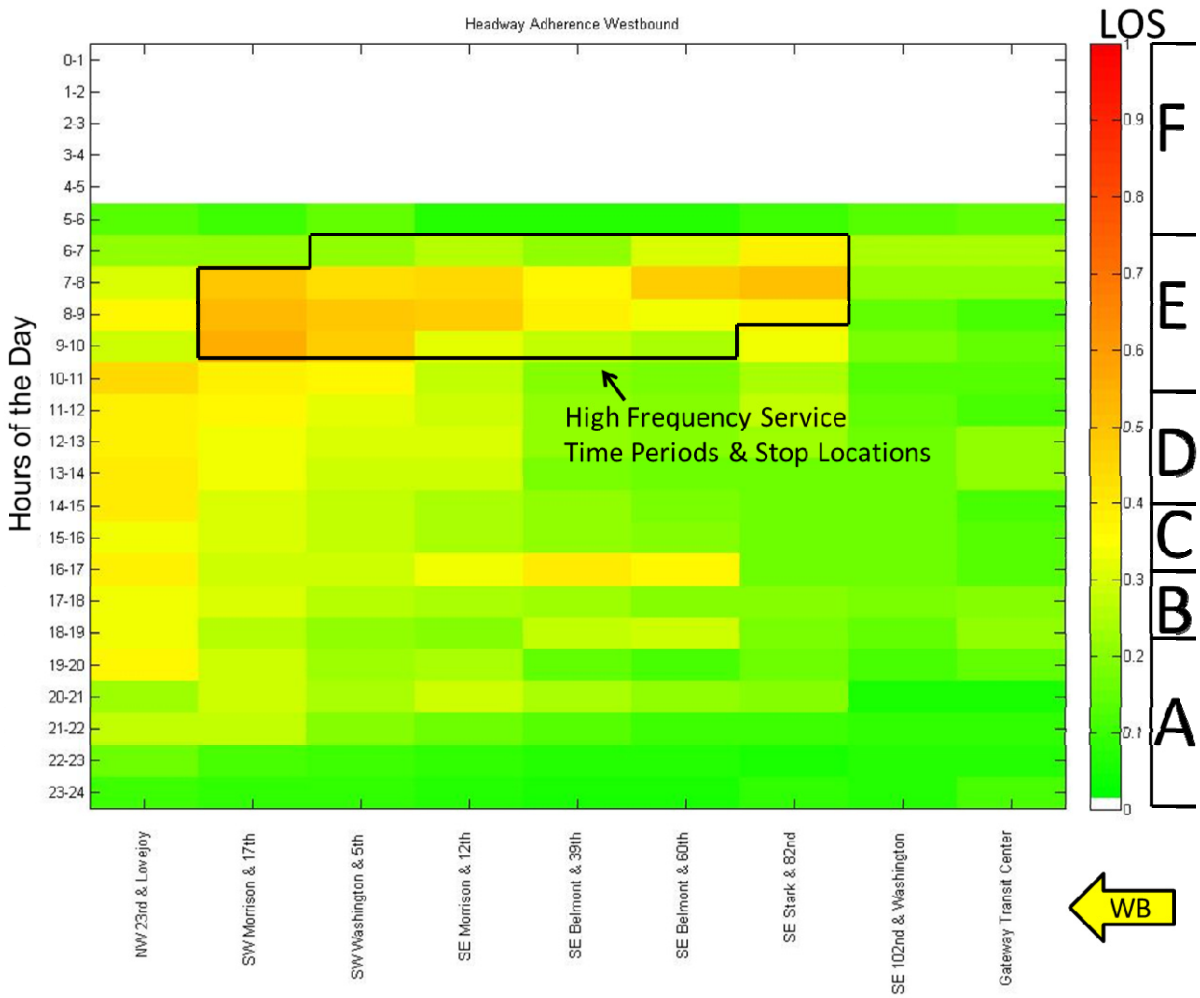


Figure 4-8. Headway adherence for Route 15 westbound

Similar to the previous contour diagram, Figure 4-8 presents the headway adherence of nine time points along the westbound (right to left) Route 15 calculated hourly. The headway adherence is again depicted by a color range – red to green, representing low LOS to high LOS. The headway adherence values and their corresponding colors are shown on the color bar. Again, the white space in the diagram means there is no data. The outlined area in the diagram indicates the high-frequency service time periods and stops; thus, the cells of interest lie within the outlined area.

Figure 4-8 suggests that, in general, headway adherence performs at mediocre LOS levels (Grade D or below) for the duration of the high-frequency service periods and for most time points. It is significant to note that this substandard performance happens immediately at the start of the high-frequency service at the SE Stark & 82nd stop. Moreover, the stop at SW Morrison & 17th appears to chronically underperform for the duration of its high-frequency service period, at LOS Grade E or worse.

Again, the contour diagram easily highlights specific problematic areas and time periods with low LOS grades for headway adherence. Inspecting the stop at SW Washington & 5th again, this time during high-frequency service between 8-9 a.m., we see it is, at best, performing at an LOS Grade D for headway adherence. This suggests that headways are not consistent for this stop and there are large deviations from the scheduled headway. Figure 4-9 and 4-10 show the running speeds (this speed is different from the Figure 4-6 travel speeds in that this running speed excluded the dwell times for stops between each two scheduled stops) and passenger boardings (these numbers of boardings are the sum of number of boardings between each two time points and include the downstream time point) for every segment in the westbound route 15 between 8-9 a.m.

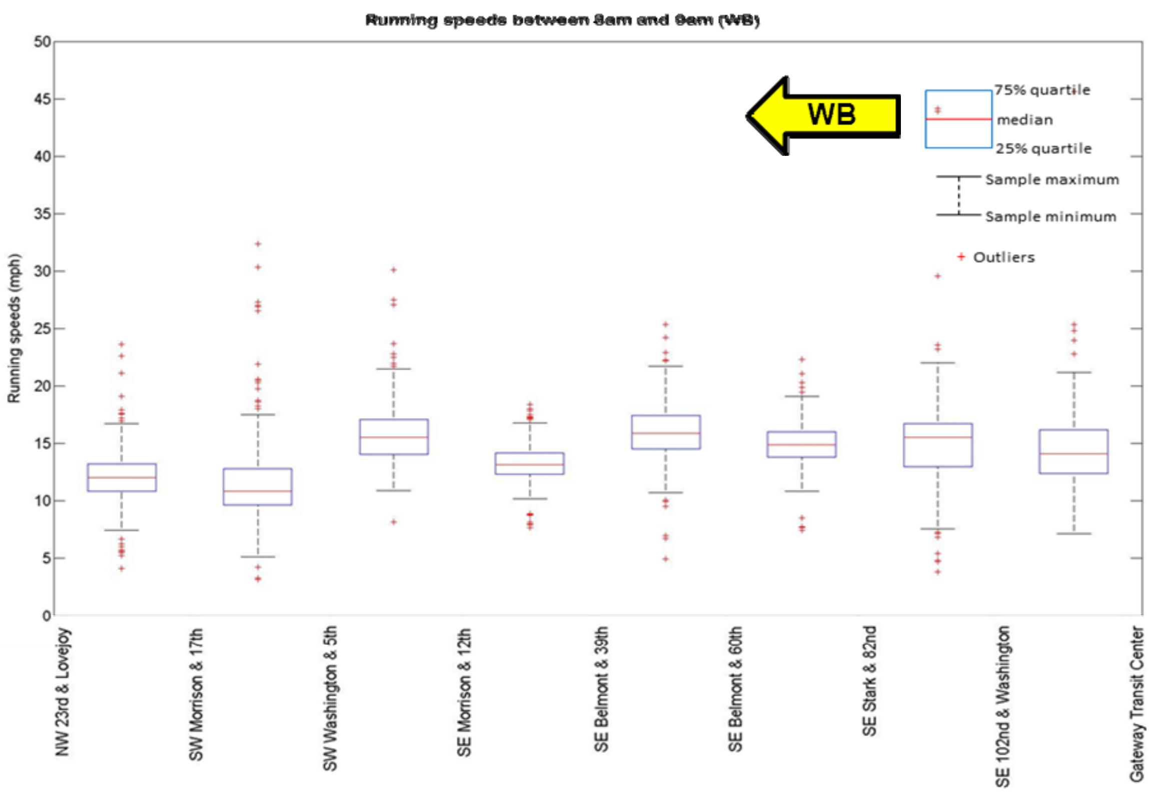


Figure 4-9. Box-plots of running speeds for Route 15 westbound

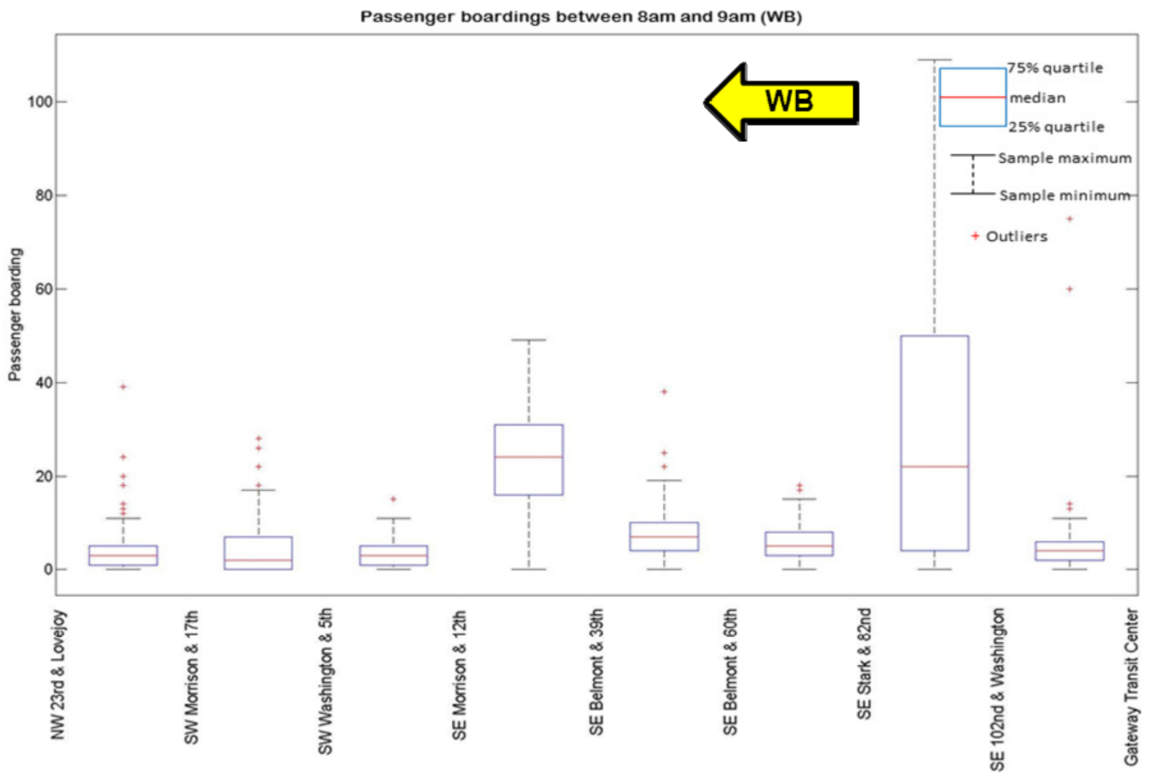


Figure 4-10. Box-plots of passenger boardings for Route 15 westbound

Figure 4-9 shows that the running speed between the SE Washington & 102nd and SE Stark & 82nd stops is relatively high. Figure 4-10 shows a large average passenger boarding and variance between these two stops at this time compared to all other segments. Therefore, the poor headway adherence for the SE Stark & 82nd stop between 8-9 a.m. shown in Figure 4-8 is mainly due to the large amount and uncertainty of passenger demand. In the next two segments, both running speeds are relatively high and passenger boarding amounts and variances are small. Therefore, the headway adherences for stops at SE Belmont & 60th and SE Belmont & 39th are comparatively better. However, after the SE Belmont & 39th stop the running speed is very slow compared to previous segments, while the passenger boarding amount and variance are much higher than other segments. Consequently, the headway adherence at SE Morrison & 12th between 8-9 a.m. suddenly becomes worse. After this effect ceases, passenger boardings become low and stable. Running speed, however, first increases in the subsequent segment and then decreases to very low values in the last two segments in the downtown Portland area. By comparing figures 4-8, 4-9 and 4-10, we can find that the segment between the SE Belmont & 39th and SE Morrison & 12th stops is the problematic one that leads to downstream poor headway adherence, which is caused by slow running speeds combined with a large amount and variance of passenger boarding activities.

Again, the color contour diagram easily highlights specific problem areas and time periods with low LOS grades for headway adherence. Inspecting the stop at SW Washington & 5th again, but this time between 8-9 a.m. (during high-frequency service), we see it is, at best, performing at an LOS Grade D for headway adherence. This suggests that headways are not consistent for this particular stop and there are large deviations from the scheduled headway. Plotting the distribution of the headway deviations records confirms this. Figure 4-11 depicts this distribution for the stop at SW Washington & 5th between 8-9 a.m.

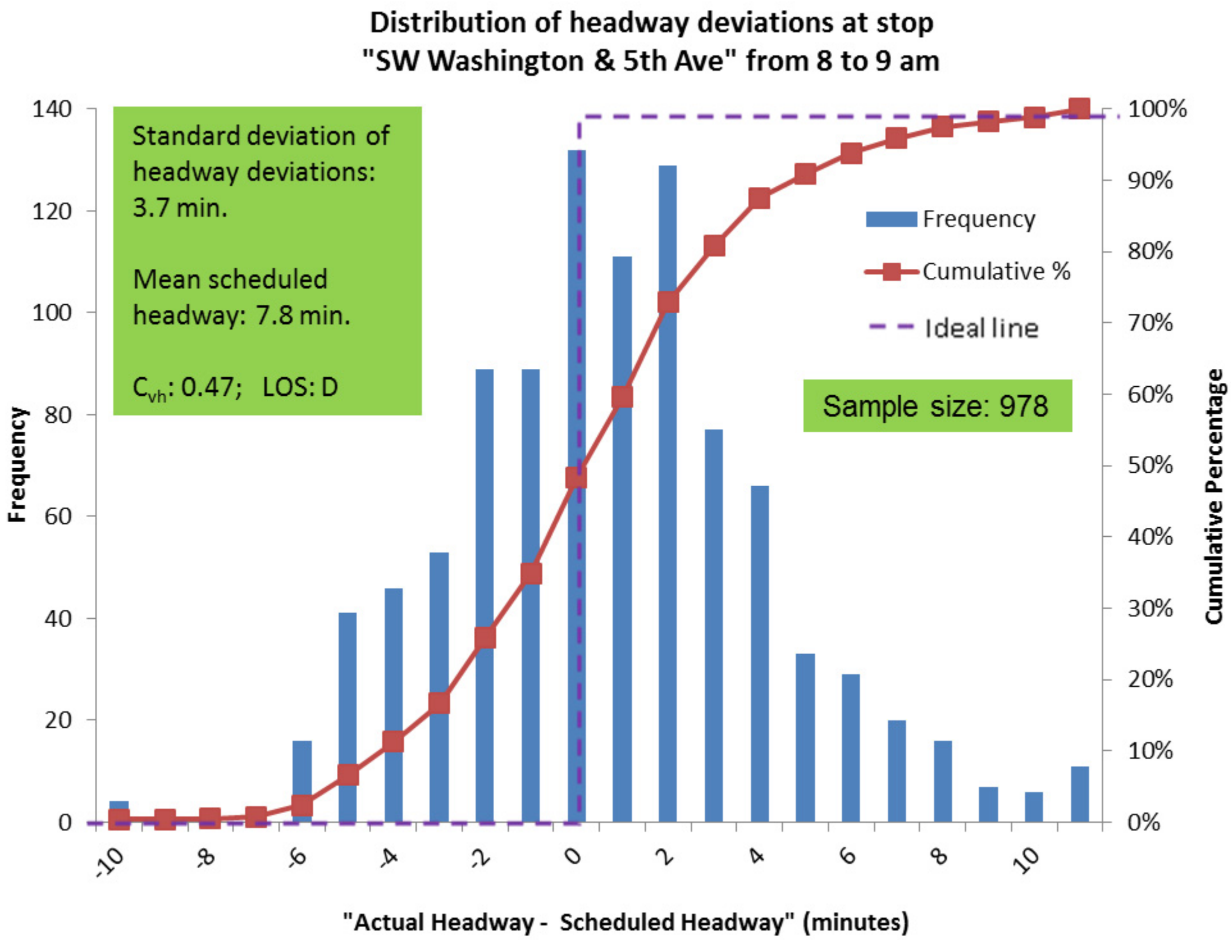


Figure 4-11. Headway adherence distribution for SW Washington & 5th stop between 9 and 10 a.m.

The dashed line in Figure 4-11 represents the ideal cumulative distribution line; the curved red line is the actual cumulative distribution line; and the bars denote how often the actual headway is greater than or less than the scheduled headway, and by how many minutes. Hence, bars to the right of the ideal line indicate actual headways were greater than scheduled headways (late arrival) and bars to the left of the ideal line denote actual headways that were less than scheduled headways (early arrival). The standard deviation of headway deviation is 3.7 minutes, which is quite high relative to the mean scheduled headway of 7.8 minutes. This suggests that if we assume the mean scheduled headway of 7.8 minutes as the scheduled headway, then passengers can expect headways anywhere from 4.1 minutes to 11.5 minutes (7.8 ± 3.7). This much variance in headways equates to a low LOS Grade D. More acceptable headway variances would have standard deviation values of 2.3 minutes (for LOS Grade B) or less.

Again, similar plots and analysis for eastbound Route 15 are not shown for brevity.

4.3 SUMMARY

This section presented a visualization framework with an interactive interface which can:

1) convert bus operational records into visible outputs (time-space diagram) and dynamically for more details (a web-based application on Google Maps); and,

2) generate and visualize reliability performance measures that can be disaggregated into temporal and spatial two-dimensions, in any one dimension (either spatial or temporal), or for a single stop in a single hour.

This research was built on previous studies and introduced the aptitude for using TriMet's archived data to create visualization tools that allow analysis disaggregated down to the hourly and stop level. The visualization techniques presented in this paper help improve the generation and display of performance measures in the following ways:

- Well-developed visualized performance measures are inherently easier to understand compared to a tabular output by drawing attention to specific stops or time periods of interest.
- When compared to other visualization studies, the data visualization techniques demonstrated here not only utilize spatial analysis along a route map, but also show how temporal analysis is possible for different hours of day – providing more thorough comprehension of how performance propagates over time and distance.
- Offering the flexibility to change the temporal aggregation level – from hourly, to AM-peak, PM-peak, and off-peak – can also provide detailed performance measurement for single-stop analysis at a selected time period.
- Generating the different visualization diagrams and graphs are automatic, accurate and applicable to other routes.
- Demonstrating the value of dynamic visualization for route monitoring by animating bus operations in order to see how their performance propagates temporally and spatially.

5.0 BUS BUNCHING ANALYSIS

Section 4.2 briefly introduced a prototype visualization tool to display bus route data. To fully understand the characteristics of unreliability, a different type of data analysis and performance evaluation is needed. This section will focus on high-frequency service unreliability characteristics, particularly bus bunching. In high-frequency service, since scheduled headways between buses are short and passenger demand is relatively high, buses more easily become bunched. A late bus usually encounters higher passenger demand and the extra passengers may create further delays. Meanwhile, if the following bus encounters fewer passengers, it tends to run faster. If the two buses become too close, "bus bunching" takes place. Bus bunching is associated with longer waiting times for most riders, uneven passenger distribution, overcrowding in late buses, and an overall decrease of level of service and capacity. Therefore, it is important for transit agencies to understand the characteristics of bus bunching so as to propose corrective strategies.

5.1 BUS BUNCHING IDENTIFICATION AND DISTRIBUTION

In this section, we propose several methodologies to identify bus bunching recurrence temporal and spatial distributions, identify bus bunching trips formation and dissipation spatially, and plot headway spatial distributions.

5.1.1 Bus Bunching Identification and Distribution

The proposed bus bunching identification method calculates the number of bus bunching events that occurred at each stop and for every hour of the day. Headway between consecutive buses at each stop is used to judge bus bunching instead of the relative distance between consecutive buses. The output is a bus bunching frequency distribution, over hours of day (or aggregated over a.m. peak hours) and stops along a route. The methodology is described in the following steps:

1) **Data cleaning.** Eliminate erroneous or mismatch records.
2) **Data sorting.** Sort by service date and leave time (actual departure time) for each stop for departure headway calculation.
3) **Data aggregation.** There might be multiple records (rows in database) for one bus that serves one stop (e.g., multiple door open and close activities). For such records, use the earliest arrival time and latest departure time, sum number of boarding and alighting passengers, and average number of onboard passengers.
4) **Headway calculation and bus identification.** Based on the above steps, headway can be easily obtained by subtracting two consecutive departure times for each stop for each day. The front bus and following bus are labeled for each headway.
5) **Bus bunching identification.** There is no clear standard minimum headway threshold to define bus bunching. An arbitrary value of three minutes is used, which means if a departure headway is smaller than three minutes, this headway is identified as a bus bunching event.

Depending on the transit agency's needs, this threshold can be modified to conduct different analyses.

6) **Bus bunching distribution.** Count the number of identified bus bunching records for each stop in each hour in the six months.

Applying this method to TriMet's westbound Route 15 and using a three-minute threshold for defining bus bunching, output is shown in the following spatial-temporal contour in Figure 5-1.

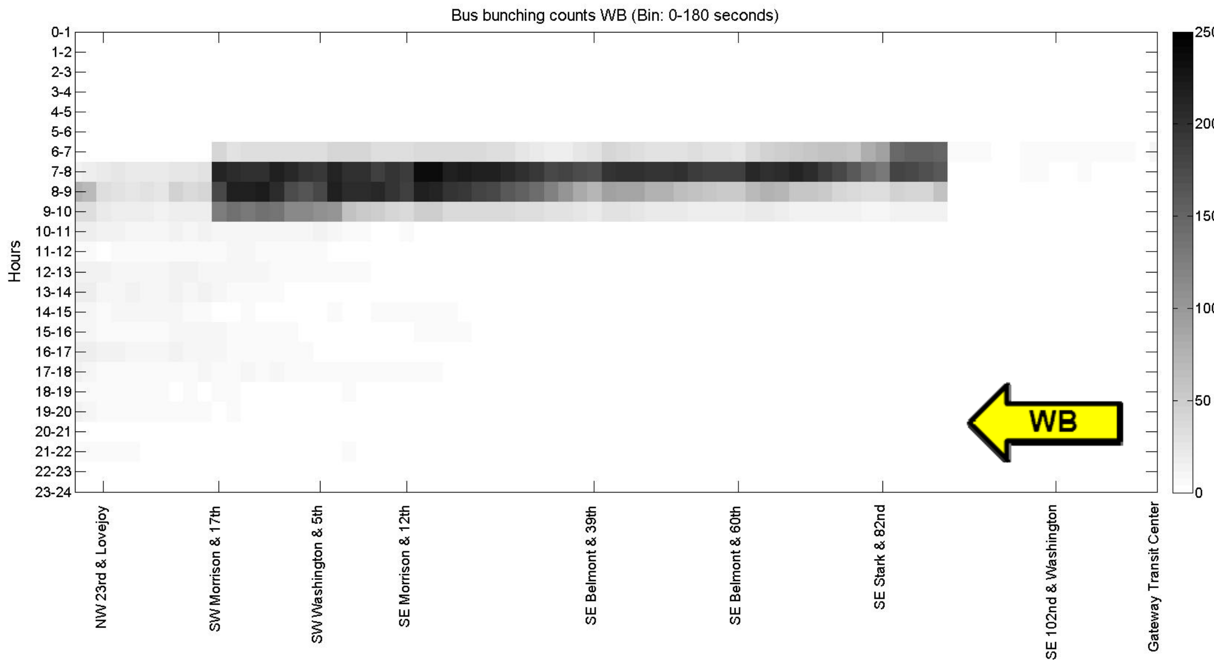
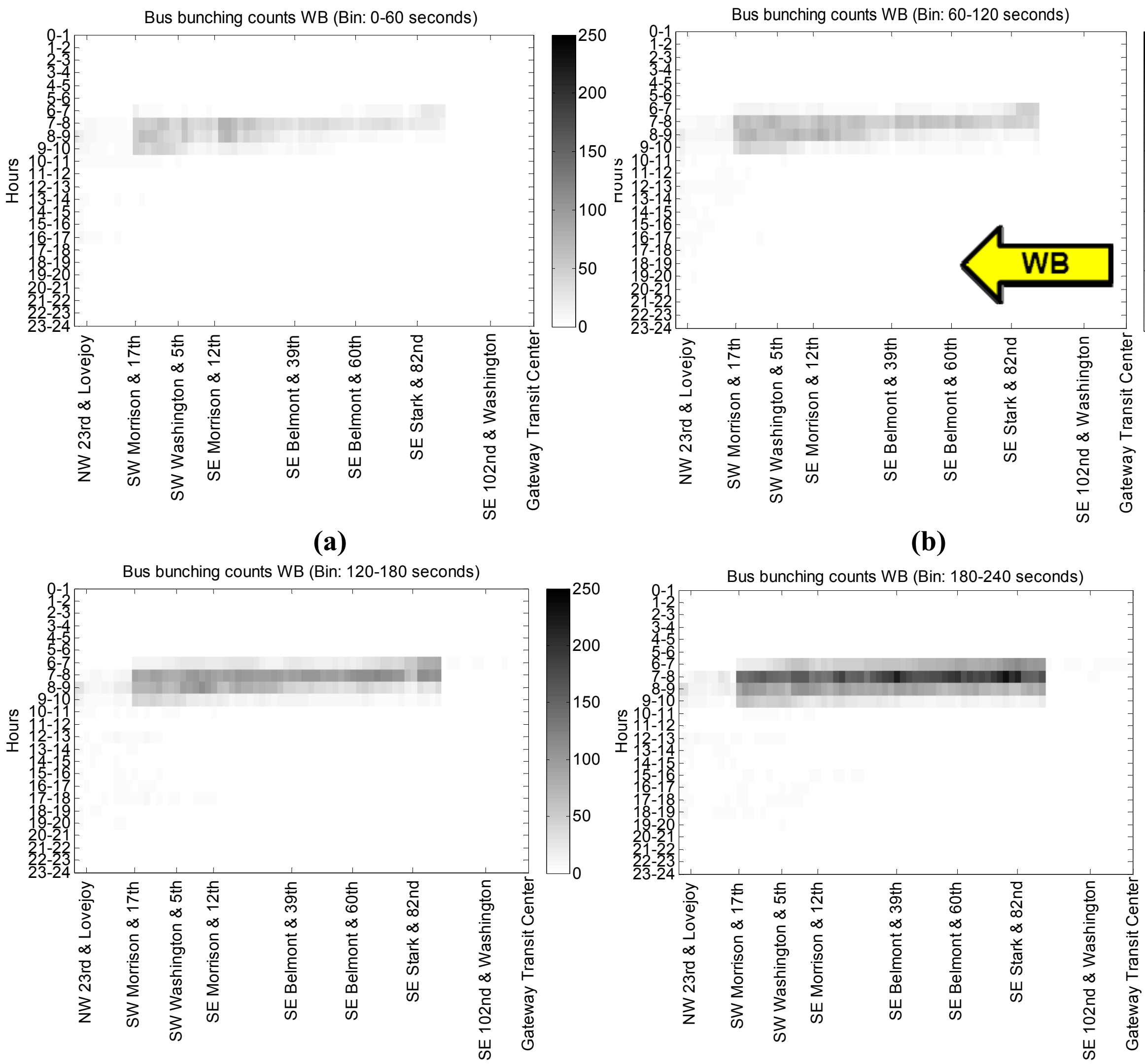


Figure 5-1. Bus bunching counts for Route 15 westbound (headway < 3 minutes)

With the exception of the two branches of stops at the western-most end of the route, the x-axis shows all scheduled stops along westbound Route 15. The y-axis shows the hours of a day. Colors from white to black represent bus bunching counts of 0 to 250 for the six-month's worth of data. Recall that the studied data set consists of 115 days; therefore, a value of 230 counts indicates that for a particular stop during a particular hour there are, on average, two bus bunching events (230/115) occurring every day. Figure 5-1 shows that most of the bus bunching problems were found within the high-frequency service (between 6-10 a.m., and between the stops at SE Stark & 93rd and SW Morrison & 17th) where average headway is about seven minutes. Also, there appears to be excessive bus bunching taking place even during the low-frequency service period after SW Washington & 5th, emphasizing poor headway regularity. Furthermore, during the high-frequency service period, the sequence of gradual color changes between each two scheduled stops suggests that the frequency of bus bunching increases after each scheduled stop and then decreases suddenly at the subsequent scheduled stop. This indicates some control strategies, such as schedule-based holding, might be taking place during these time periods and at these stops. Figure 5-1 shows the results using three minutes (180 seconds) as the bus bunching threshold. It is also worthwhile to show how bus bunching frequency responds to different thresholds.

Figure 5-2 shows the bus bunching frequency distribution for four different short-headway threshold levels, again over time and space. The color bar and relative values are the same as in Figure 5-1. The general behavior to note is that as the short-headway threshold increases, the bus bunching frequency also increases. Further, most of the identified short headways are occurring during the high-frequency service period.



(a) (b) (c) (d)

Figure 5-2. Bus bunching counts for Route 15 westbound for thresholds (a) 0-1 minute, (b) 1-2 minutes, (c) 2-3 minutes, (d) 3-4 minutes.

The above mentioned method can generate all of the identified bus bunching events, while when two buses are bunched together at some stops along a route they will probably continue to bunch together downstream. Therefore, it is also worth identifying when and where buses started to bunch together. Figure 5-3 shows the number of bus bunching initials.

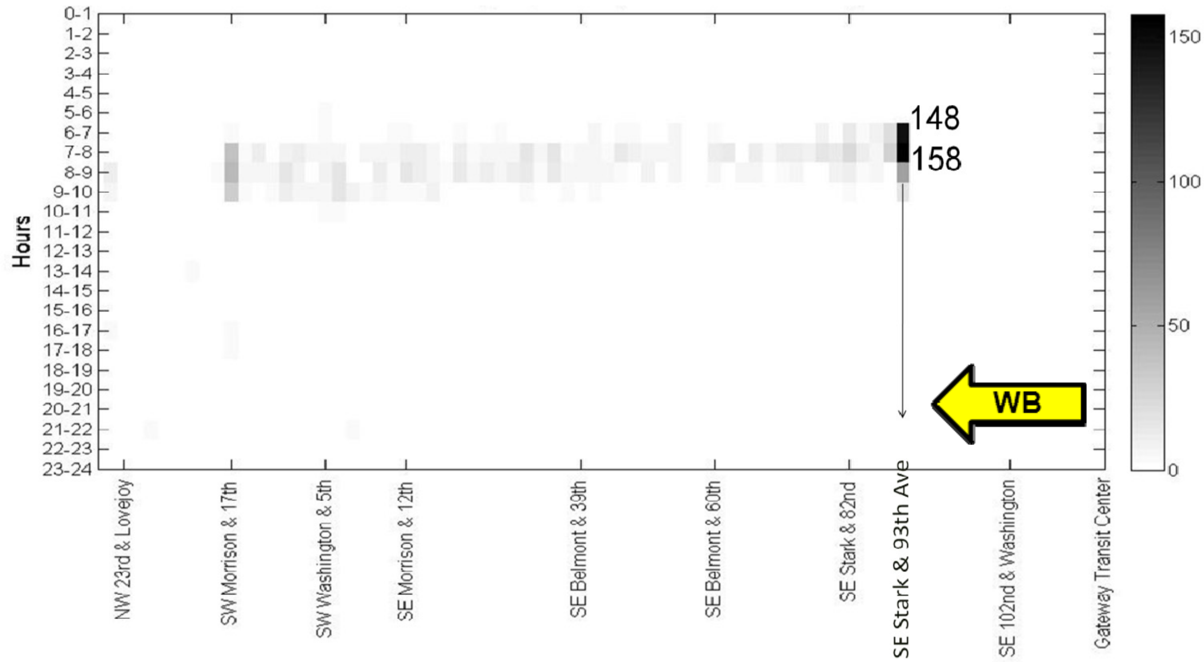


Figure 5-3. Bus bunching initials for Route 15 westbound (headway threshold: < 3 minutes)

As shown in Figure 5-3, most of the bunching patterns start from the same stop - SE Stark & 93rd - in the morning peak hours. The numbers 148 and 158 are the total counts during the 115 weekdays; therefore, on average, there is more than one bus bunching trip start at this stop every weekday every hour between 6-8 a.m. This stop is the first stop for the high-frequency service segment along westbound, which means additional buses join in the westbound trip at this stop during the morning peak hour. There are basically two potential reasons for this. First, according to the current TriMet schedule-based operating strategy, even if a bus upstream at the SE Stark & 93rd stop is late, buses that are scheduled to start service from this stop will still be on time. Second, bus operators may not start the trip on time if they arrived back to the garage from an opposite trip late because there are labor laws dictating rest time for bus operators.

5.1.2 Headway Delay and Actual Headway Spatial Distributions

Section 5.1.1 identifies the bus bunching counts defined by departure headway at any stop using predefined thresholds. Large gaps distributions and unusual headways are also necessary to analyze because they have significantly negative impacts on passenger waiting times. Therefore, we also provide another two outputs that show headway delay and actual headway distributions, as shown in Figure 5-4 and Figure 5-5.

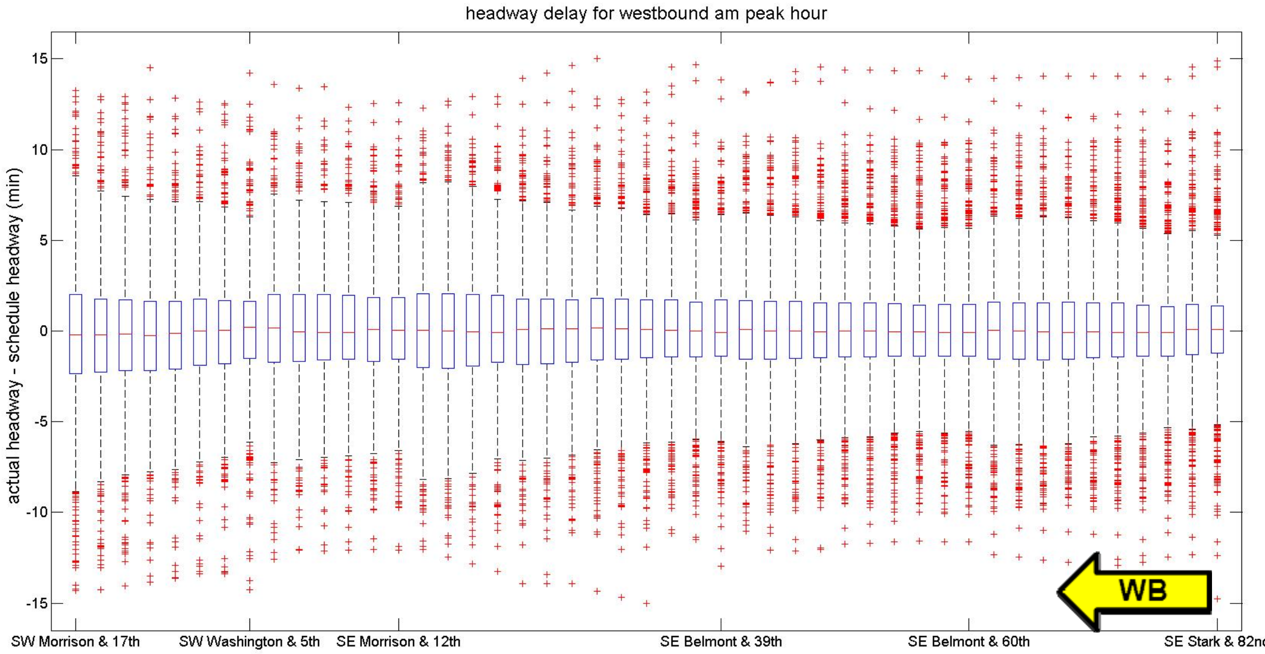


Figure 5-4. Headway delay spatial distribution for Route 15 westbound a.m. peak hours

Figure 5-4 shows the headway delay (actual headway – schedule headway) distribution for all westbound stops in the high-frequency service zone in morning peak hours; 50% of headway delays (blue box for each stop) are within two minutes for all westbound stops. The boundaries of headway delay records (dashed line outside the blue box at each stop) grow gradually towards the east end from ± five minutes to ± eight minutes, and decrease after each time point. This indicates that the longer the distance, the more there's a possibility of large headway delay while a time point helps maintain regular headways.

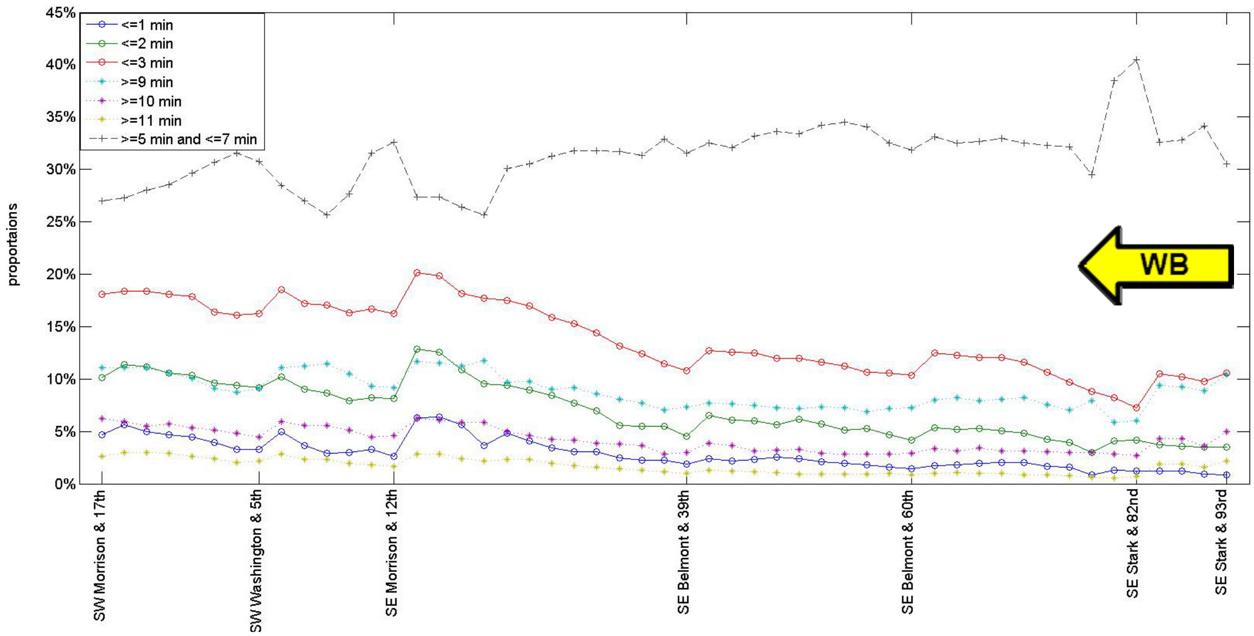


Figure 5-5. Actual headway proportions spatial distribution for Route 15 westbound a.m. peak hours

Figure 5-5 shows the proportions of actual headways in different bins for all westbound stops in the high-frequency time and zone. Around 25%-35% of the actual headways are within the scheduled headway boundary (five to seven minutes) for all stops except the stop at SE Stark & 82nd (40%). The proportion has a decreasing trend towards the east end with a mild increase at each time point. This figure also shows the proportion change for all other levels of irregular headways, which have a general increasing trend towards the east end and a mild decrease at each time point.

5.1.3 Bus Bunching Trips Formation and Dissipation

Section 5.1.1 shows spatial and temporal distribution of bus bunching counts defined by the departure headway at each stop; however, these bus bunching events are identified independently without knowing which trips they belong to. It is worth knowing where buses get bunched and separated for all pairs of bus trips. Therefore, we developed another method to identify bus bunching trips and where buses get bunched and separated for each pair of trips. The basic idea of this algorithm can be described as follows:

1) **Data cleaning.** Eliminate erroneous or mismatch records.
2) **Data aggregation.** There might be multiple records (rows in database) for one bus that serves one stop (e.g., multiple door open and close activities). For such records, use the earliest arrival time and latest departure time, sum the number of boarding and alighting passengers, and average the number of onboard passengers.
3) **Data sorting.** Sort by service date and trip number at the SE Stark & 93rd stop because, in the high-frequency service, some trips start from this stop in addition to the terminal at the Gateway Transit Center.
4) **Headway calculation.** For each pair of trips, create two lists that store their departure time at each stop along the route, and calculate the departure headway at each matched

stop. If the headway is smaller than an input bunching threshold (e.g. < two minutes), flag that stop.

5) **Plot.** Create a matrix for each pair of bus trips. If at least one stop is flagged as bunching, add that trip to the matrix with all calculated departure headways, and highlight all those flagged stops.

By applying this method, Figure 5-6 shows the results of bus bunching trips formation and dissipation for Route 15 westbound a.m. peak hours. In this figure, each horizontal line represents an identified bus bunching trip in which at least one stop has a departure headway that is smaller than the bunching threshold. In each horizontal line, blue dots represent bunching records at corresponding stops. If consecutive stops in one pair of trips are identified bunching, blue lines connect them.

From Figure 5-6, we can visualize where bus bunching trips form and dissipate most frequently. It is very obvious that most of the bus bunching trips start at the SE Stark & 93rd stop because this stop is the start of additional trips. If any trip starting from the west terminal arrives at this stop late, while an additional short trip starts from this stop on time, there might be bus bunching here. Similarly, most of the dissipation of bus bunching trips happens at the SW Morrison & 17th stop. One can also visualize how the bus bunching density increases to the west until the stop at SW Morrison & 17th.

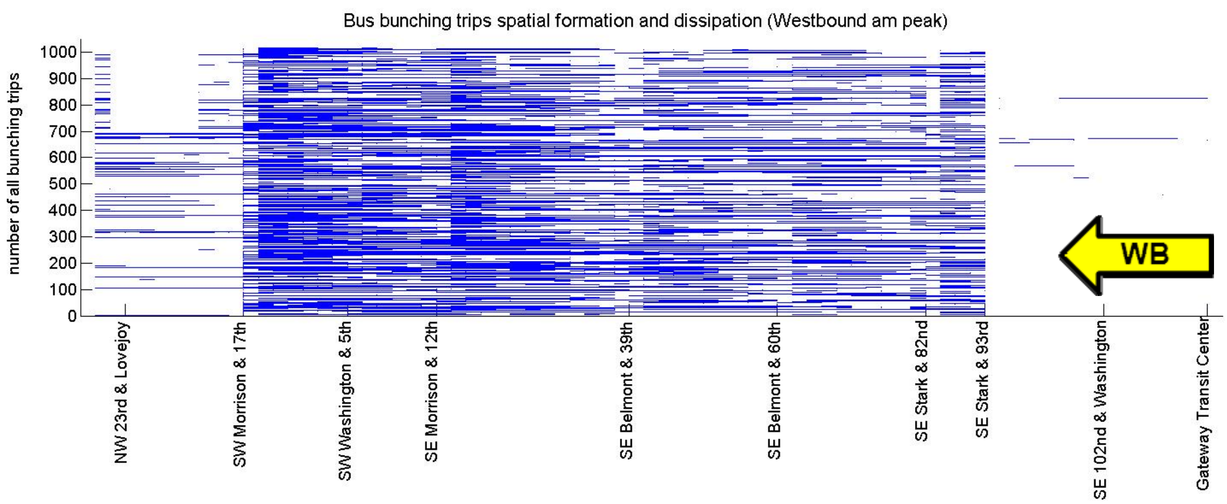


Figure 5-6. Bus bunching trips spatial formation and dissipation for Route 15 westbound a.m. peak hours (headway threshold: < 2 minutes)

While Figure 5-6 shows the formation and dissipation of all bus bunching trips, it is also worth knowing where they first formed and where they finally dissipated. Figure 5-7 further shows the percentage of first formation and last dissipation of all bus bunching trips at each stop along the route, so that particular control strategies can be implemented at those stops. The blue bars represent the percentages of formation, and red bars represent the percentages of dissipation. Figure 5-7 shows that almost 15% of all the westbound a.m. peak bus bunching trips first formed at the SE Stark & 93rd stop, with all the other stops less than 5%. On the other hand, almost 20% and 25% of all the bus bunching trips dissipated at SW Morrison & 17th and its downstream stop, totaling 45%. This observation also indicates that once a pair of buses get bunched, they are more likely to continue until the end of high-frequency service zone. Therefore, a better headway-based control at this stop in the morning peak hours is highly recommended.

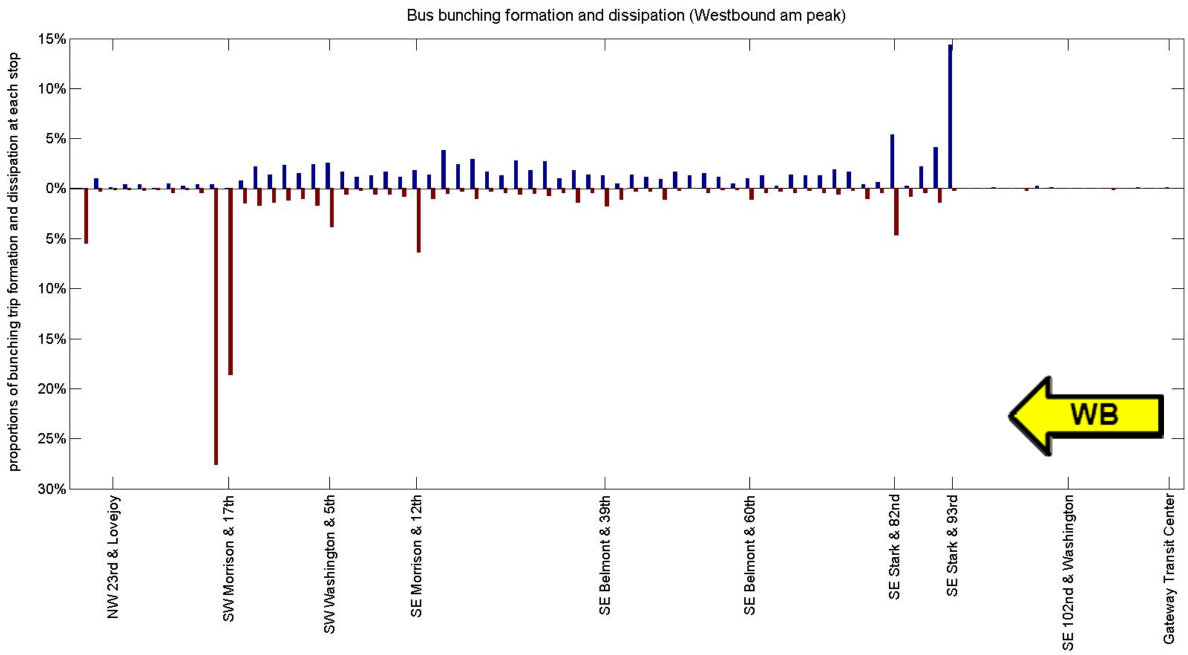


Figure 5-7. Bus bunching trips first formation and last dissipation spatial distribution for Route 15 westbound a.m. peak hours (headway threshold: < 2 minutes)

Figure 5-8 further shows the probabilities of downstream bus bunching trips at different departure headway bins at the SE Stark & 93rd stop for westbound a.m. peak hours, with three different headway threshold levels. For example, if bunching is defined as headway threshold < one minute, and departure headway at the trip beginning is in the zero- to one-minute bin, the probability of a downstream bus bunching trip is 70% (blue line); if the departure headway at the beginning trip is in the four- to five-minutes bin, the probability of a downstream bus bunching trip is less than 10%. Similar interpretations apply for the green line (headway threshold < two minutes) and red line (headway threshold < three minutes). The general trends for all three lines are decreasing as the departure headway bins at the beginning trip increases, which indicates that the longer headway at the starting trip, the larger probability of downstream bus bunching trips.

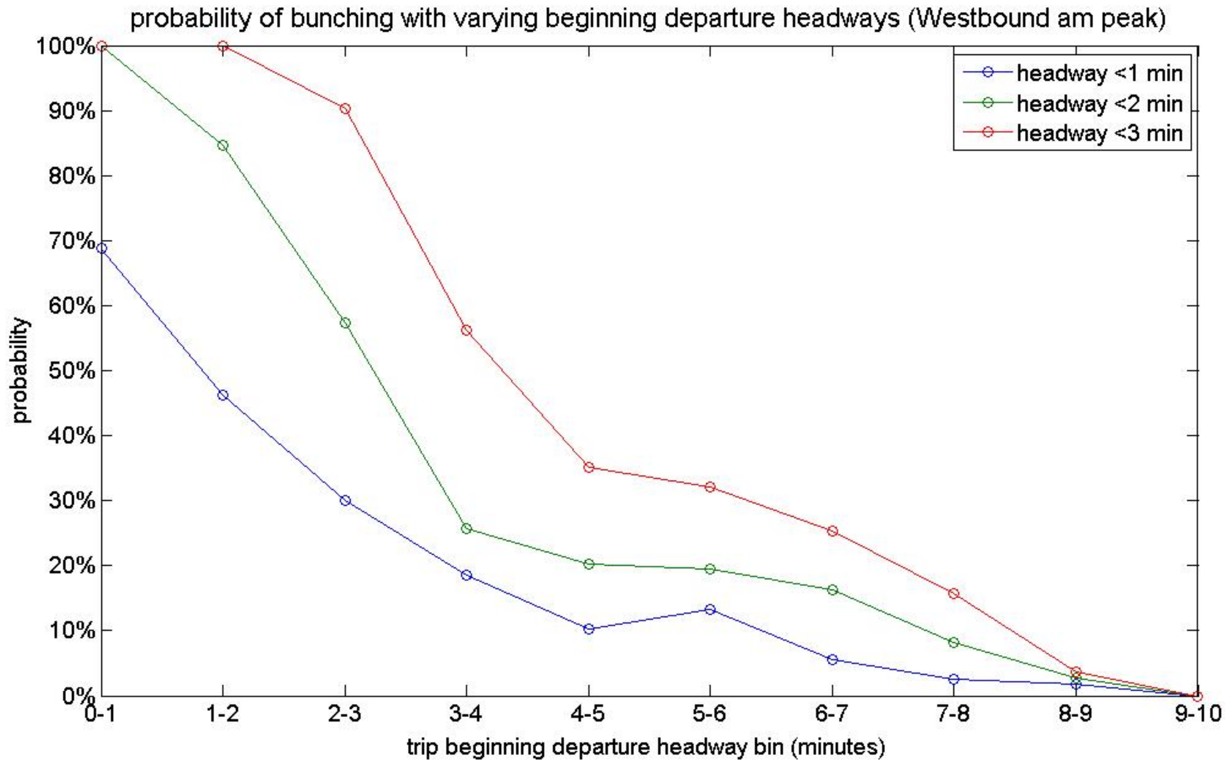


Figure 5-8. Probability of downstream bunching with varying trip beginning departure headways and headway thresholds for Route 15 westbound a.m. peak hours

5.2 BUS BUNCHING EFFECTS AND POTENTIAL CAUSES

Section 5.1 illustrated the methodologies of evaluating headway performance, identifying bus bunching characteristics, and ways of reporting these results visually. However, what are the impacts of bus bunching or irregular headways on passengers? What factors cause bus bunching? The first question is important because it determines whether it is necessary for transit agencies to take any action to mitigate bus bunching occurrence, and the second question is important because it helps transit agencies choose the right strategies to improve service reliability.

Section 5.2 first evaluates the effects of bus bunching on passenger waiting time and overcrowding, and then analyzes multiple factors that are related to bus bunching and potential causes in the stop-level and segment-level.

5.2.1 Bus Bunching Effects

Bus bunching has several negative impacts on passengers, such as longer waiting times and overcrowding. In a pair of bunched buses, the leading bus is usually late and the following bus is usually earlier than its scheduled departure times. Therefore, passengers who wait for the leading bus may have to wait longer than a scheduled headway at any stop, while passengers who arrive at a stop after the leading bus departure and before the following bus arrival may have a very short waiting time. However, the number of passengers who wait for the leading bus is usually more than those passengers wait for the following bus, and therefore may result in a weighted average waiting time for passengers that are involved in a pair of bunching trips longer than normal average waiting time. To evaluate if this is true, Figure 5-9 shows the average times for normal buses, leading buses, following buses, and weighted average waiting time for both leading and following buses in a pair of bunching trips. It shows that waiting time for the leading bus is almost 1.5 minutes more than the normal bus, and almost four minutes more than the following bus in an environment where scheduled headway is between five and seven minutes. Also, the weighted average waiting time for both the leading and following buses in a bunching trip is about a half-minute more than the normal buses on average over all stops. It seems like not too much, but if you assume the average penalty of passenger waiting time is $20/hour, 10% bunching trips, and passenger demand is 500 people per hour, the annual cost will be over $3,000 only for this route and this direction in the morning peak hours, not to mention the afternoon peak hours and other routes, and that potential ridership may decrease.

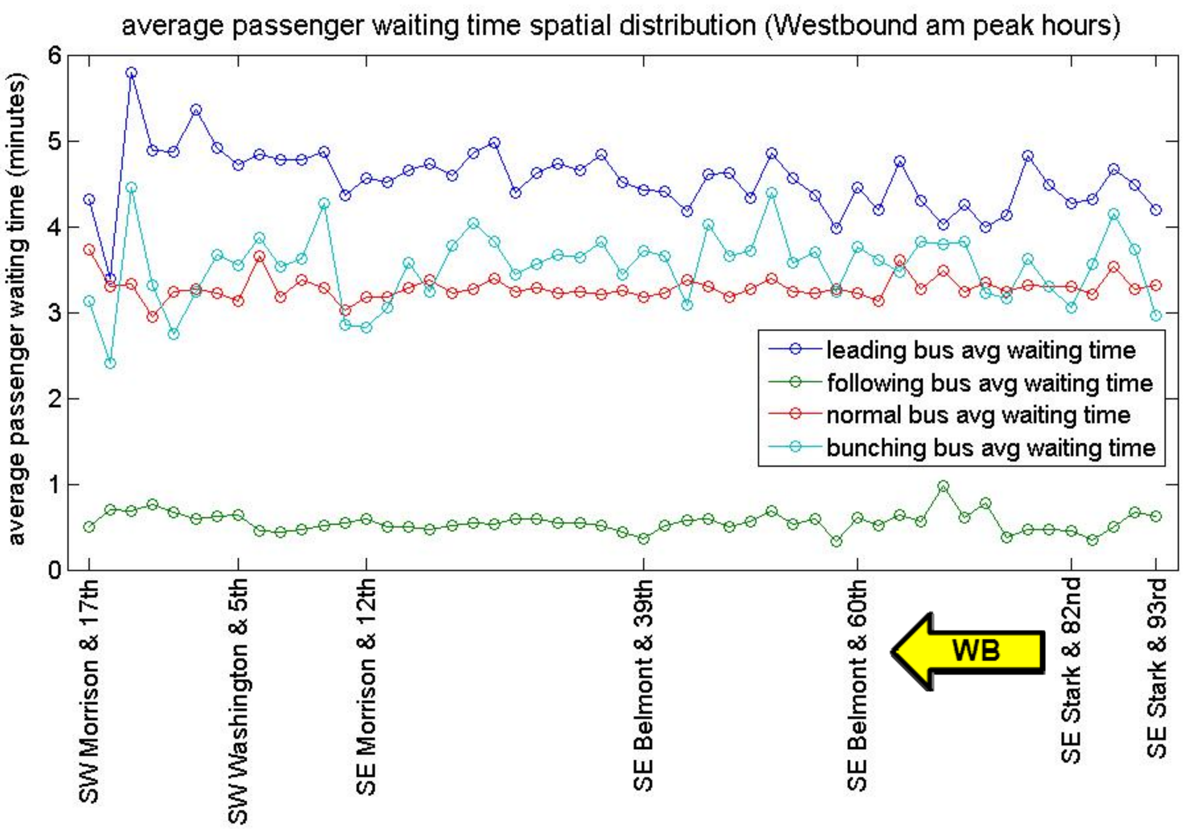


Figure 5-9. Average passenger waiting time spatial distribution for Route 15 westbound a.m. peak hours (headway threshold: < 2 minutes)

Figure 5-10 shows that the average passenger load on a leading bus between the stops at SE Belmont & 39th and SW Washington & 5th is almost 10 people more than on a normal bus, or a weighted average of bunching buses, and 20 more than that on a following bus. For all other stops, the difference is not significant due to low passenger demand. Although the average load on a normal bus is almost the same as the weighted average load on a pair of bunching buses, it still indicates overcrowding in the leading bus and low occupancy in the following bus, which results in a loss of attractiveness to passengers and reduction in bus utilization efficiency.

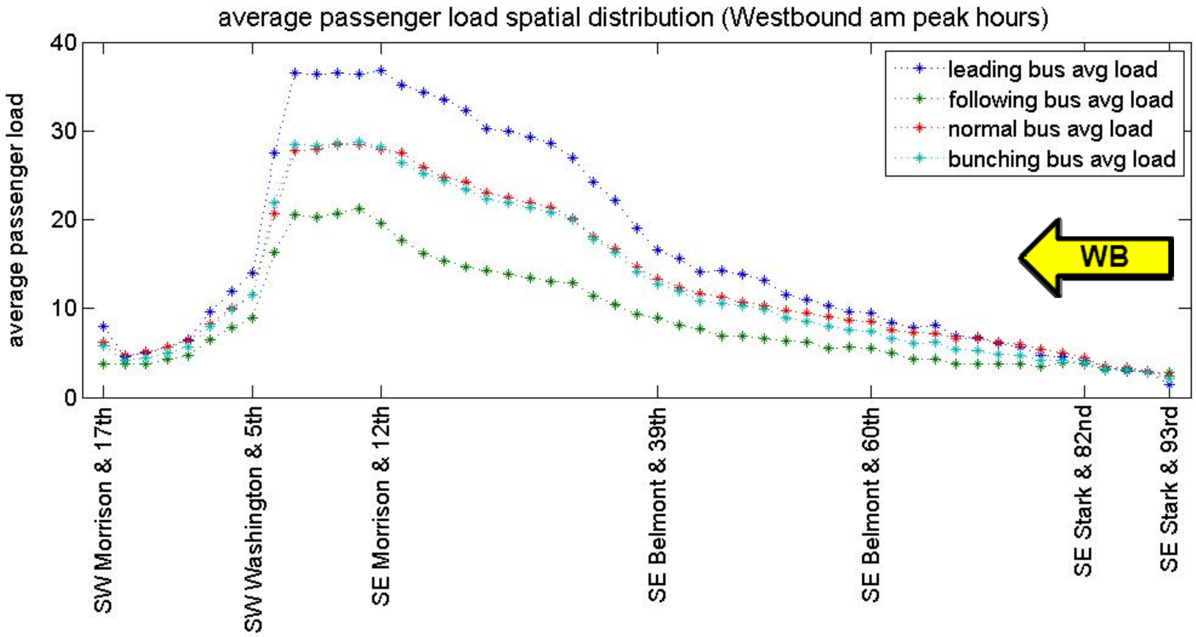


Figure 5-10. Average passenger load spatial distribution for Route 15 westbound a.m. peak hours (headway threshold: < 2 minutes)

5.2.2 Bus Bunching Causes Analysis

Section 5.2.1 has shown the negative impacts of bus bunching on passenger waiting time and overcrowdings. Next, it is necessary to understand some of the attributes and real causes of bus bunching.

Stop-level analysis

In this section, we introduce a methodology that summarizes the attributes of bus bunching at the stop-level and provides sensitivity analysis of some thresholds used in the method.

Based on our definition of bus bunching – where departure headway at a stop is less than a determined threshold – the cause of such bus bunching must be due to either a late departure for the front bus, an early departure for the following bus, or both. Whenever bus bunching occurs, one or a combination of the following attributes is taking place:

- Front bus - late departure from previous stop;
- Front bus - long travel time from previous stop to current stop;
- Front bus - long dwell time or "stay time" (departure time minus arrival time) at current stop;
- Following bus - early departure from previous stop;
- Following bus - short travel time from previous stop to current stop;
- Following bus - short dwell time or stay time at current stop.

Due to the complexity, a set of exclusive combinations for all six attributes is not provided in this study; rather, the frequency of each attribute that is associated to an identified bus bunching record for all records is calculated.

Factors that affect travel time may include traffic congestion, signal timing, accident/incident, bus operator behavior, and the number of on-board passengers. Factors that affect dwell time, or “stay time,” may include passenger boarding and alighting movement, lift use, other buses at the same stop, or the stop location relative to a signalized intersection. As this research is based on analyzing archived BDS stop-event data, only a portion of these potential factors are calculated for this study. Table 2 shows the attributes that are analyzed in this study:

Table 5-1. Bus Bunching Attributes Summary

Front bus	***Following bus***
1. Late departure from previous stop	1. Early departure from previous stop
2. Long travel time from previous stop to current stop	2. Short travel time from previous stop to current stop
3. Long dwell time at current stop	3. Short dwell time at current stop
4. Long “stay time” at current stop	4. Short “stay time” at current stop
5. Large passenger movement (boarding + alighting)	5. Small passenger movement (boarding + alighting)
6. High passenger load	6. Low passenger load
7. Lift use	

The objective for this section is to calculate how often (using percentages) each attribute meets its threshold for all identified bus bunching records. Thresholds are pre-determined for all the attributes. The process is illustrated below in six major steps:

1) For each hour of the day and for each stop, compute the average travel time, dwell time, “stay time,” passenger movement, and passenger load over the studied six months’ BDS data. These mean-value matrices will determine appropriate thresholds for each measure;
2) Load the bus bunching results, where the front bus and following bus are identified for each record;
3) Calculate the schedule adherence for both buses at their previous stops;
4) Calculate the travel time from the previous stop to the current stop for both buses;
5) Calculate dwell times, stay times, passenger movements, passenger loads and lift uses for both buses at the current stop;
6) Compare each of the measures calculated between Step 3 and Step 5 to the mean-value matrices that were calculated in Step 1. Calculate the frequency of how often each measure (which represents each of the attributes) meets their thresholds. The default thresholds used for this study are arbitrarily determined and can be modified:
    a. Late departure > two minutes;
    b. Early departure < minus one minute;
    c. Lift use > one;
    d. All other long/heavy thresholds > 1.5 times the mean value;

e. Short/light thresholds < half of the mean value;

Results and sensitivity analyses are summarized in the following tables: Table 5-2 shows the variability under different headway levels, Table 5-3 shows the difference under different threshold sets, and Table 5-4 shows the sensitivity of segment.

Table 5-2. Bus Bunching Attributes Statistics (Route 15 Westbound)

			Headway levels (minutes)				
			0-3	0-1	1-2	2-3	3-4
	Bus Bunching Attribute	Thresholds	%	%	%	%	%
Front bus	Late departs from previous stop	>2 min.	70%	80%	79%	59%	33%
	Long dwell time at current stop	>1.5*mean	33%	36%	34%	31%	28%
	Long stay time at current stop	>1.5*mean	26%	27%	26%	25%	24%
	Large passenger movement at current stop	>1.5*mean	40%	42%	42%	38%	33%
	High passenger load at current stop	>1.5*mean	40%	41%	42%	38%	29%
	Lift use	> = 1	0%	0%	0%	0%	0%
	Long travel time from previous stop to current stop	>1.5*mean	11%	12%	11%	10%	9%
Following bus	Early depart from previous stop	< -1 min.	32%	27%	33%	34%	28%
	Short dwell time at current stop	< .5*mean	67%	67%	67%	67%	64%
	Short stay time at current stop	< .5*mean	29%	29%	30%	29%	27%
	Small passenger movement at current stop	< .5*mean	70%	69%	70%	71%	66%
	Low passenger load at current stop	< .5*mean	42%	40%	43%	41%	34%
	Short travel time from previous stop to current stop	< .5*mean	8%	9%	8%	7%	8%

Table 5-2 summarizes statistics of common bus bunching attributes. For this study, each bus bunching record is related to the attributes listed in first major column of Table 5-2. This column describes different causes and consequences for either the front bus or the following bus. In order to categorize each characteristic related to bus bunching, thresholds were defined. Each characteristic can be categorized (i.e., either "late" or "not late," "long" or "not long," etc.). The thresholds column provides limits to define how long is "long" or how late is "late." For example, the front bus characteristic "departs from previous stop" is considered "late" when the departure is "later than two minutes." The column for headway levels is broken down into five sub-columns for different short headway levels: 1) zero to three minutes, which is the default short headway level for this study; and four other incremental levels: 2) zero to one minute, 3) one to two minutes, 4) two to three minutes, and 5) three to four minutes. The percentage value shows how frequently each characteristic meets its threshold for all the bus bunching records.

For example, for the short headway level of zero to three minutes and the "Front bus" with the characteristic "departs from previous stop," the value is 70%. This signifies that for all the bus bunching records identified with headway level zero to three minutes, 70% of front bus departures from the previous stop are later than two minutes. Looking at another result for the same short headway level of zero to three minutes for the "Following bus" with the characteristic "dwell time at current stop," the value is 67%. This means that for all the bus bunching records identified with headway level zero to three minutes, 67% of following bus dwell times at the current stop are shorter than half of the average dwell times (for that stop).

Table 5-2 reveals the three attributes that are most frequently related to bus bunching, based on their high percentage values: 1) Front bus – “departure from previous stop,” 2) Following bus – “dwell time at current stop,” and 3) Following bus – “passenger movement at current stop.” Note the percentage of “lift use at current stop” for the front bus is very small for all short headway levels (0% due to round). Also, the travel time attributes for either the front bus or following bus have relatively small percentages. The front bus travel times are less likely to be 1.5 times more than the average travel time (all headway levels are less than 12%). The following bus travel times are less likely to be less than half the average travel time (all headway levels are less than 9%). This implies that most of the short headways are not due to travel time variation, but due to poor departure adherence and dwell time variation.

The results shown in Table 5-2 are affected by the threshold definitions; however, since the thresholds are arbitrary it is valuable to test for threshold sensitivity. Table 5-3 shows the percentages of each attribute for three different threshold categories – classified as “easy,” “moderate,” or “difficult” – all the results are calculated using the same short headway level (zero to three minutes).

Table 5-3. Attributes Sensitivity Analysis (Headway Threshold: < 3 Minutes, WB)

		Easy		Moderate		Difficult	
		Threshold		Threshold		Threshold	
Front bus	Departs from previous stop	> 1 min.	84%	> 2 min.	70%	> 3 min.	50%
	Dwell time at current stop	> mean	47%	> 1.5*mean	33%	> 2 times	23%
	Stay time at current stop	> mean	52%	> 1.5*mean	26%	> 2 times	12%
	Passenger movement at current stop	> mean	50%	> 1.5*mean	40%	> 2 times	31%
	Passenger load at current stop	> mean	59%	> 1.5*mean	40%	> 2 times	21%
	Lift use	>= 1	0%	> = 1	0%	> = 1	0%
	Travel time from previous stop to current stop	> mean	52%	> 1.5*mean	11%	> 2 times	5%
Following bus	Departs from previous stop	< 0 min.	66%	< -1 min.	32%	< -2 min.	11%
	Dwell time at current stop	< .5*mean	67%	< .33*mean	63%	< .25*mean	62%
	Stay time at current stop	< .5*mean	29%	< .33*mean	13%	< .25*mean	7%
	Passenger movement at current stop	< .5*mean	70%	< .33*mean	65%	< .25*mean	65%
	Passenger load at current stop	< .5*mean	42%	< .33*mean	28%	< .25*mean	23%
	Travel time from previous stop to current stop	< .5*mean	8%	< .33*mean	4%	< .25*mean	3%

In Table 5-3, note new threshold definitions for each of the three categories which are easy, moderate or difficult to meet. Note that with the exception of lift use, all of the results show that as the thresholds become more difficult to meet, the attribute percentages decrease. However, some attributes prove to be less sensitive to the threshold categories. For example, the percentages for the “Following bus” attribute of “dwell time at current stop” are 67%, 63% and 62% for the easy, moderate and difficult thresholds, respectively. This suggests that almost all of the following bus dwell time records, which are less than half of the average value, are also less than a quarter the average value. A similar trend is found for another following bus attribute, “passenger movement at current stop.” This is logical since passenger movement is a major contributor to dwell time. Another point of interest for this analysis is the difference in percentage values for the following bus attributes of "dwell time at current stop" (67%, 63% and 62%) and "stay time at current stop" (29%, 13% and 7%). Ideally, these two attributes should have similar percentage values across all threshold categories, since a bus should be departing shortly after serving passengers. However, these results indicate that many of the following buses wait at a current stop after serving passengers.

Finally, to determine whether the attribute percentage statistics vary for different segments along the route, it is necessary to compare percentages for different route segments. Table 5-4 illustrates the results for five segments along westbound Route 15. Again, the short headway level remains at zero to three minutes and the default values are used for attribute thresholds.

Surprisingly, Table 5-4 shows relatively large differences between the percentages across the five route segments. For example, for stops in segment between "SE Belmont & 60th and SE Belmont & 39th", 79% of all the bus bunching records are correlated to the following bus's dwell time at the current stop being less than half of the average value; while for this same attribute, it accounts for only 59% of bus bunching records for stops in the segment between "SE Morrison & 12th and SW Washington & 5th". The difference of attribute percentages from one segment to another is significant. This level of analysis may prove extremely valuable for uncovering the source of bus bunching incidents by drilling down to stop-level analysis to reveal peculiar performance behavior.

Table 5-4. Spatial Comparisons of Attributes Statistics (Headway Threshold: < 3 Minutes, WB)

			Segment %				
		Thresholds	SW Morrison & 17th/ SW Washington & 5th	SW Washington & 5th/ SE Morrison & 12th	SE Morrison & 12th/ SE Belmont & 39th	SE Belmont & 39th/ SE Belmont & 60th	SE Belmont & 60th/ SE Stark & 82nd
Front bus	Departures from previous stop	> 2 min.	76%	76%	68%	63%	66%
	Dwell time at current stop	> 1.5*mean	26%	23%	39%	35%	32%
	Stay time at current stop	> 1.5*mean	21%	15%	30%	30%	28%
	Passenger movement at current stop	> 1.5*mean	42%	30%	48%	41%	33%
	Passenger load at current stop	> 1.5*mean	20%	62%	60%	41%	23%
	Lift use	> = 1	0%	0%	0%	0%	0%
	Travel time from previous stop to current stop	> 1.5*mean	17%	10%	7%	4%	7%
Following bus	Departure from previous stop	< -1 min.	29%	31%	43%	33%	15%
	Dwell time at current stop	< .5*mean	52%	59%	62%	79%	86%
	Stay time at current stop	< .5*mean	18%	30%	34%	32%	19%
	Passenger movement at current stop	< .5*mean	54%	65%	66%	81%	87%
	Passenger load at current stop	< .5*mean	54%	23%	30%	38%	53%
	Travel time from previous stop to current stop	< .5*mean	14%	4%	5%	7%	4%

Segment-level analysis

The above section analyzed some potential causes and consequences of bus bunching at the stop level. While given the fact that headway changes between two consecutive stops are usually too small, it may not be easy to conclude convincing relationships between attributes and bus bunching events. It may be interesting to analyze from the segment level. We propose another algorithm that summarizes bus bunching occurrence probability at a downstream time point as a result of departure headway level at the upstream time point, and various potential contributing factors. The result is shown in Figure 5-11.

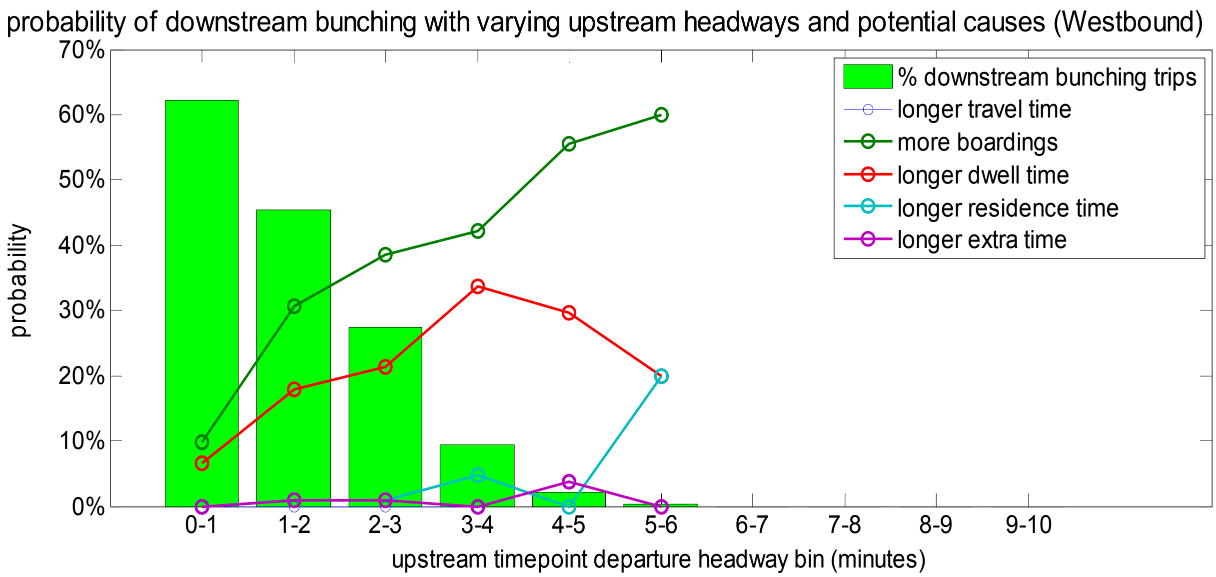


Figure 5-11. Probability of downstream bunching with varying upstream headways and potential causes for Route 15 westbound a.m. peak (headway threshold: < 2 minutes)

The x-axis lists all the possible departure headways at any upstream time point, and the green bars represent the summarized probabilities of downstream time point bus bunching occurrence. For example, in the upstream time point departure headway bin two to three minutes, the green bar indicates that of all the departures at upstream time points, if the departure headway between two buses is two to three minutes, on average, around 30% of the arrival headways (or departure headway) at their downstream time points, will be less than the threshold (< two minutes, various sensitivity analysis are tested but not shown here for brevity). Therefore, these green bars are decreasing as the upstream time point departure headway bin grows, which means the shorter the departure headway at the upstream time point, the larger probability of getting bunch at downstream time point.

The values of the five lines mean the percentages of trips where these factor variables are statistically significant is different between leading buses and following buses in trips that are defined as bunching at the downstream time point. Take the green line, for example, in the upstream time point departure headway bin three to four minutes: 10% of them get bunched at the downstream time point (green bar), and again in 40% of these trips, passenger boardings for the leading bus are statistically significant more than the number of boardings for the following bus at stops within a segment. The green line in the figure is increasing as the upstream time point departure headway bin grows, which indicates that as upstream time point departure headways increase, significantly more boardings in the leading bus than the following bus plays a more important role on those trips that get bunched at downstream time points. Note that "residence time" means the actual departure time – actual arrival time at any stop, and "extra time" means "residence time" – dwell time at any stop. The large difference between the dwell time line and residence line indicate a number of holding controls might have been applied to the following bus. The line for travel time is always zero, which means travel times between each two stops (exclude all at-stop times, pure travel time) within a segment is always not significantly different. In other words, the main causes for buses from not bunching to bunching are at-stop variability, and particularly, passenger boardings, not travel time variability.

The algorithm of generating this result can be briefly described as below:

1) Sort trips by service_date and trip_number, so that in each day, the trip starting times at the first stop of high-frequency service zone are in increasing order.
2) For each pair of trips;
    For each segment along the high-frequency service zone;
        For each bin of upstream time point departure headway;
            i. calculate the arrival headway (or departure headway) at the downstream time point, and flag if it is smaller than the input threshold;
            ii. create five arrays for all five factor variables with two columns in each array and store the leading and following buses activities at each stop in that segment excluding the two time points (including the downstream time point activity if the downstream departure headway is used as the threshold);
            iii. do a one-way ANOVA analysis for each array and flag if data in the leading bus column are significantly larger than that in the following bus column.
        End
    End
   End
3) For each bin of upstream time point departure headway;
    a. calculate the percentage of downstream time point arrivals that are headway smaller than bunching threshold (bunching);
    b. calculate the percentage of significantly different trips that are flagged as bunching at a downstream time point for all factor variables.
   End

5.3 SUMMARY

In this section, we have shown various methods that identify bus bunching characteristics, and evaluate the effects and causes of bus bunching quantitatively. Some important results and findings can be summarized as below:

1) Bus bunching usually takes place in the high-frequency service time and zone, mainly because the scheduled headway is short.

2) Bus bunching trips usually start at the first stop of a high-frequency service zone; there is no headway control at this stop and bus operators cannot communicate with each other.
3) When departure headway at the first stop of a high-frequency service zone is less than four minutes, there is a high probability of bus bunching worsening downstream.
4) Passenger waiting time for a leading bunched bus is 1.5 minutes longer than for a normal (not bunched) bus, and four minutes more than for a bunched following bus; weighted average passenger waiting time for pair of bunching buses is 0.5 minutes longer than normal buses, on average.
5) Bus loads on leading buses, is on average, 10 extra passengers more than normal (not bunched) bus and 20 extra passengers more than a bunched following bus.
6) Late departure from the last stop for the leading bus and less passenger boardings for the following bus are the two factors that are more associated with bus bunching events.

6.0 REGRESSION MODELS

Section 5 presented techniques to identify bus bunching, spatial and temporal distribution; understand characteristics; evaluate effects; and analyze potential causes. Although these numerical analyses provide plenty of information, statistical models that relate some of the key variables with various factors can provide a better understanding of the key determinants of service reliability variables and to what extent they affect the reliability variables.

This section presents three regression models for dwell time, travel time, and headway, respectively, as explained by a set of explanatory variables.

6.1 DWELL TIME

The key factors that affect dwell time are passenger boarding and alighting and fare payments. Also, as explained in section 5.2, at-stop activity variability is the main cause of bus bunching. Numerous dwell time models can be found in the literature; a comprehensive summary can be found in (Liao C., 2011). However, dwell time regression models are usually not applied universally because the fare payment system, bus design, and many other factors may differ in different transit systems. Also, the dwell time model developed for TriMet by Dueker, Kimpel & Strathman (2004) needs an update since the system has been changing over the past six years. While the data used in Dueker, Kimpel & Strathman are from various types of routes, this study utilizes data only from Route 15 as the first step.

Since we have a very large sample of data, data cleaning is necessary before the model estimation. We removed dwell times longer than three minutes and dwell times at time points to avoid the potential impact of holding.

The estimated model results are shown in Table 6-1. Dependent variable "dwell time" is measured in seconds. Results show that all variables are significant. The average time for an individual door opening, passenger boarding, alighting and lift use are 2,225, 3.35, 1.475 and 37.799 seconds, respectively, holding all other variables at their mean. Dwell time for a low-floor bus is 0.825 seconds less than a high-floor bus, on average, holding other variables at their mean.

Compared to the Dueker, Kimpel & Strathman model, we add door-open times as a new variable, with 2.225 seconds on average and a very high t-value. Coefficients for all other variables are similar except lift use, which is 38 seconds in this model compared to 62 seconds in their model. The adjusted R-square 0.57 is also higher than their model, which is 0.34.

Table 6-1. Dwell Time (seconds) Regression Model

Variable names	*Coefficients*	*Std. of Coeff.*	*t-value*
Constant	1.51	0.045	33.5
# of door openings	2.225	0.016	139.1
# of boardings	3.350	0.010	324.1
# of boardings square	-0.014	0.001	-10.0
# of alightings	1.475	0.009	168.4
# of alightings square	-0.018	0.001	-23.5
# of lift uses	37.799	0.080	473.8
Low-floor bus dummy	-0.825	0.078	-10.6
Adjusted R-square		0.57	
Sample size		509,360	

6.2 TRAVEL TIME

In this section, two travel time regression models are presented: travel time and pure travel time. Both are estimated in the segment level, which means travel time data records are collected by "arrival time at a downstream time point – departure time at an upstream time point." Pure travel time data records are calculated by travel time in a segment minus the sum of dwell times at all stops within that travel time.

Table 6-2. Travel Time (minutes) Regression Model

Variable names	*Coefficients*	*Std. of Coeff.*	*t-value*
Constants	1.300	0.048	26.9
Distance (miles)	2.213	0.013	170.7
# of stops	0.279	0.004	75.3
# of signalized intersections "through"	0.075	0.002	30.4
# of signalized intersections "right turn"	0.331	0.008	41.1
# of signalized intersections "left turn"	0.633	0.11	59.6
# of stop signs	0.209	0.010	20.6
Schedule delay at upstream time point (min)	-0.016	0.002	-9.7
Headway delay at upstream time point (min)	-0.006	0.001	-9.9
Sum of boardings	0.059	0.003	46.9
Sum of alightings	0.020	0.001	13.8
Sum of lift uses	0.626	0.022	27.9
A.m. peak inbound dummy	0.004	0.015	0.27
P.m. peak outbound dummy	0.497	0.018	28.3
Low-floor bus dummy	-0.617	0.043	-14.5
Adjusted R-square		0.472	
Sample size		104,010	

Before estimation, we first generated all the travel time records and with all the matched explanatory variables. The travel time variables here are in the segment level between each two time points or terminal stops. The distance, number of signalized intersections, and stop signs associated with each segment are obtained manually from a TriMet interactive map. The sum of boardings, alightings and lift uses are for those stops in each segment, but exclude the two time points. We also had to eliminate errors after the data preparation process. We removed those travel times records that experienced speeds higher than 40 mph or speeds lower than five mph in any segment, and we also removed records with negative travel times, unexpected long or short distances, or empty headway delays (mainly due to wrong GPS data). After all these process, the results are summarized in Table 6-2 and Table 6-3.

Results indicate that, on average, each additional mile adds 2.213 minutes to travel time. The additional time for passing through a signalized intersection (0.075 minutes, 4.5 seconds) is much shorter than making a turn at a signalized intersection, and a left turn takes much longer (0.633 minutes, 38 seconds) than a right turn (0.331 minutes, 20 seconds) at a signalized intersection, on average. The additional times for an additional stop (0.279 minutes, 17 seconds) or a stop sign (0.209 minutes, 13 seconds) are close, which mainly represents the deceleration and acceleration delay. One minute of departure schedule delay from an upstream time point results in a 0.016 minutes (10 seconds) decrease in travel time, and one minute of departure headway delay from upstream time point results in a 0.006 minutes (four seconds) decrease in average travel time. These two variables indicate significant but small reductions of travel time due to schedule or headway delay from an upstream time point. Each additional number of boardings, alightings or lift uses at any stop will add 0.059, 0.02 and 0.626 minutes (3.54, 1.2 and 37.56 seconds) to travel time, which are similar to the coefficients estimated in the dwell time model. If a bus is traveling westbound in the a.m. peak hours there is no significant difference in travel time, but if a bus is traveling eastbound in the p.m. peak hours an average of 0.497 minutes (29.82 seconds) additional time is added to travel time. Also, a low-floor bus results in an average of 0.617 minutes (37.02 seconds) of significant reduction in travel time.

Table 6-3. Pure Travel Time (minutes) Regression Model

Variable names	*Coefficients*	*Std. of Coeff.*	*t-value*
Constants	1.510	0.045	33.5
Distance (miles)	2.203	0.120	182.6
# of stops	0.186	0.002	75.4
# of signalized intersections “through”	0.057	0.002	25.4
# of signalized intersections “right turn”	0.442	0.007	59.1
# of signalized intersections “left turn”	0.469	0.010	47.7
# of stop signs	0.050	0.009	5.3
Schedule delay at upstream time point (min)	-0.016	0.002	-10.2
Headway delay at upstream time point (min)	-0.006	0.001	-9.9
A.m. peak inbound dummy	0.112	0.014	7.8
P.m. peak outbound dummy	0.312	0.016	19.4
Low-floor bus dummy	-0.572	0.040	-14.4
Adjusted R-square		0.410	
Sample size		104,010	

Because pure travel time records exclude dwell times at stops in any segment, compared to the travel time model, three explanatory variables “sum of boardings,” “sum of alightings” and “sum of lift uses” are removed. Most of the coefficients are close to those in the travel time model.

However, some of the difference needs to be highlighted: Compared to travel time models, the coefficients for making a right turn and left turn in this model are very close (0.442 and 0.469 minutes), which indicate similar extra time is added to travel time for right or left turns at signalized intersections. The additional time for a stop sign in this model is very short at only 0.05 minutes (3 seconds), compared to 13 seconds in the travel time model. An additional 0.112 minutes (6.72 seconds) will be added to the pure travel time if a bus is traveling westbound in the a.m. peak hours, and in this model it is significant but not in the travel time model. An additional 0.312 minutes (18.72 seconds) will be added to the pure travel time if a bus is traveling eastbound in the p.m. peak hours, which is much shorter than the travel time model (29.82 seconds), but both are significant.

6.3 HEADWAY

In this section, we present two headway regression models in which we can predict arrival or departure headways at a time point based on departure information from the upstream time point and activities of two buses between two time points.

The headway records are all from time points in a high-frequency service time and zone (a.m. peak for westbound and p.m. peak for eastbound). In each headway record, two buses are distinguished as front bus (depart earlier) and back bus (depart later) at time points. So in each record in the data, there are two dependent variables, arrival headway and departure headway, and several explanatory variables that include departure headway of the same two buses from the upstream time point, and the boardings; alightings and lift use for each bus between the upstream time point and the current time point; and the holding time at the current time point. This holding time variable is calculated by the following formula:

Max [departure time at current time point – arrival time at current time point – estimated dwell time at current time point, 0]

The estimated dwell time at the current time point is calculated by applying the above estimated dwell time model with the current time point activities (boardings, alightings and lift use) for the front and back buses. Model estimation results are shown in Table 6-4. Results indicate that an average of 0.373 minutes headway at the current time point holding all other variables at their means. For each additional departure headway from the upstream time point, an average of 0.934 and 0.941 minutes, respectively, will be added to the current time point departure headway and arrival headway.

All other variables for front bus and back bus are similar in magnitude but with opposite signs. For each additional passenger boarding on the front bus, the departure and arrival headways at the current time point will have an average of 0.061 and 0.059 minutes (3.66 and 3.54 seconds) reduction, and for each additional passenger boarding on the back bus, the departure and arrival headways at the current time point will have an average of 0.063 and 0.062 minutes (3.78 and 3.72 seconds) increase. There are similar explanations for the front bus and back passenger alightings and lift uses. In the departure headway model there are another two variables, which are holding times for the front and back buses. On average, one minute holding for the front bus at the current time point leads to an average of 0.743 minutes (44.58 seconds) reduction in departure headway at current time point, and one minute holding for the back bus at the current time point leads to an average of 0.758 minutes (45.48 seconds) increase in departure headway at the current time point. Both of the models have very high R-squares: 0.903 and 0.884 for departure headway and arrival headway models, respectively.

Table 6-4. Headway Regression Models (minutes)

	Departure headway (min)			**Arrival headway (min)**		
Variable names	*Coeff.*	*Std. of Coeff.*	*t-value*	*Coeff.*	*Std. of Coeff.*	*t-value*
Constants	0.373	0.034	10.829	0.328	0.034	9.524
Departure headway at upstream time point (min)	0.934	0.003	293.123	0.941	0.004	262.934
Front bus sum of boardings	-0.061	0.002	-37.819	-0.059	0.002	-32.496
Back bus sum of boardings	0.063	0.002	39.742	0.062	0.002	34.926
Front bus sum of alightings	-0.017	0.002	-9.822	-0.016	0.002	-7.958
Back bus sum of alightings	0.018	0.002	10.438	0.018	0.002	9.193
Front bus sum of lift uses	-0.606	0.064	-9.442	-0.531	0.002	-7.327
Back bus sum of lift uses	0.700	0.064	10.994	0.595	0.002	8.274
Front bus holding time at current time point (min)	-0.743	0.015	-49.103			
Back bus holding time at current time point (min)	0.758	0.014	53.321			
Adjusted R-square	0.903			0.884		
Sample size	9,862			9,862		

6.4 SUMMARY

In this section, we present several statistical models for dwell time, travel times and headways. Model estimation results show that passenger boardings, alightings and lift use are significant in all models if they are included, and the coefficients are very stable for each one. The average effects of stops, signalized intersections and stop signs on travel time are estimated, which are not studied in detail in the literature. Travel time models also indicate a higher travel time in the p.m. peak eastbound than the a.m. peak westbound or other off-peak hours for Route 15. A low-floor bus can save an average of 37 seconds in travel time. The distance coefficient of 2.2 minutes/mile indicates the average travel speed for Route 15 is 27 mph.

7.0 CONCLUSIONS AND RECOMMENDATIONS

The ability to accurately and effectively generate and analyze operations performance measures is critical for transit agencies to ensure efficient transit operations and management.

This study used archived AVL/APC data from TriMet to develop an algorithm that identifies hourly and daily bus bunching occurrences for each stop along a bus route. The results have demonstrated how bus bunching incidents can be summarized for different temporal and spatial aggregation levels. The thresholds (defined in our study as the headway between two consecutive buses for a single stop) for identifying bus bunching are flexible and can be modified as an input variable according to the needs of various transit agencies.

In our case study of TriMet's Route 15, with the westbound direction it was observed that bus bunching (using less than three minutes as a threshold) typically happens during the high-frequency service period between 6-10 a.m. and between the stops at SE Stark & 82nd and SW Morrison & 17th. Moderate occurrences of bus bunching take place from 7 a.m. to 8 p.m. at stops located west of SE Morrison & 12th. Similar results were found when separating the zero- to three-minute bus bunching threshold into four different levels.

This study also develops a method to summarize causes of identified bus bunching incidents. Seven cause factors for the front bus and six cause factors for the following bus were analyzed. We first determined the frequency of each cause (expressed as percentages) meeting pre-determined thresholds for all bus bunching records. Next, we performed a sensitivity analysis to demonstrate how cause percentage results change using varying difficulty levels of bus bunching thresholds. Finally, we investigated how cause percentage results vary spatially along different route segments. By comparing the results from the three analyses, transit agency decision makers can gain much insight to understanding the causes of bus bunching along a route, thereby proposing confident and constructive strategies to mitigate further occurrences.

In this section, we have shown various methods that identify bus bunching characteristics, evaluate effects and provide an understanding of the causes of bus bunching quantitatively. Some important results and findings can be summarized as below:

1) Bus bunching usually takes place in the high-frequency service time and zone, mainly because the scheduled headway is short;
2) Bus bunching trips usually start at the first stop of a high-frequency service zone; there is no headway control at this stop and bus operators cannot communicate with each other;
3) When departure headway at the first stop of a high-frequency service zone is less than four minutes, there is a high probability of bus bunching worsening downstream;
4) Passenger waiting time for a leading bunched bus is 1.5 minutes longer than for a normal (not bunched) bus, and four minutes more than for a bunched following bus; weighted average passenger waiting time for a pair of bunching buses is 0.5 minutes longer than normal buses, on average;

5) Bus loads on leading buses is, on average, 10 extra passengers more than a normal (not bunched) bus and 20 extra passenger than a bunched following bus; and
6) Late departure from the last stop for the leading bus and less passenger boardings for the following bus are the two factors that are more associated with bus bunching.

Time points are helpful in regulating headways but with serious limitations; the lack of driver information transmission regarding front and following bus headway limits headway recovery if both buses are late (or early). In terms of significant factors affecting travel times:

1) Passenger boardings, alightings and lift use are significant in all models if they are included, and the coefficients are very stable in every one;
2) A higher travel time in the p.m. peak eastbound than in the a.m. peak westbound or other off-peak hours for Route 15;
3) The low-floor bus can save an average of 37 seconds in travel time; and
4) The distance coefficients 2.2 minutes/mile indicate an average travel speed for Route 15 is 27 mph.

The high impact of passenger and stop times indicate that one of the best strategies for headway recovery would be to limit boarding in late buses if the following bus is severely bunched and arriving one or two minutes later.

8.0 REFERENCES

Abkiwitz, M., Waksam, R., Englisher, L., & Wilson, N. (1978). *Transit Service Reliability.* Boston, MA: Department of Transportation: Cambridge.

Berkow, M., Chee, J., Bertini, R., & Monsere, C. (2007). *Transit Performance Measurement and Arterial Travel Time Estimation Using Archived AVL Data.* IEEE.

Berkow, M., EI-Geneidy, A. M., Bertini, R. L., & Crout, D. (2009). Beyond Generating Transit Performance Measures: Visulaizations and Statistical Analysis Using Hisotical Data. *Journal of the Transportation Research Record*, 158-168.

Bertini, R. L., & El-Geneidy, A. M. (2003). Generating Transit Performance Measures with Archived Data. *Journal of the Transportation Research Record*, 109-119.

Bertini, R., & El-Geneidy, A. (2004). Modeling Transit Trip Time Using Archived Bus Dispatch System Data. *Journal of Transportation Engineering*, 56-67.

Ceder, A. (2007). *Public Transit Planning and Operation: Theory, Modeling and Practice.* Oxford, Burlington: Elsevier.

Dueker, K., Kimpel, T., & Strathman, J. (2004). Determinants of Dwell Time. *Journal of Public Transportation*, 21-40.

El-Geneidy, A., Horning, J., & Krizek, K. (2010). Analyzing transit service reliability using detailed data from automatic vehicular locator systems. *Journal of Advanced Transportation*.

Feng, W., & Figliozzi, M. (2011). Using Archived AVL-APC Bus Data to Identify Spatial-Temporal Causes of Bus Bunching. *The 90th Annual Meeting of Transportation Research Board.* Washington, D.C.

Furth, P. (2000). *TCRP Synthesis 34: Data Analysis for Bus Planning and Monitoring.* Washington, D.C.: Transportation Research Board of National Acedemies.

Furth, P., & Muller, T. J. (2007). Service Reliability and Optimal Running Time Schedules. *Journal of Transportation Research Record*, 55-61.

Furth, P., Hemily, B., Muller, T., & Strathman, J. (2006). *TCRP Report 113: Using Archived AVL-APC Data to Improve Transit Performance and Management.* Washington, D.C.: Transportation Research Board of the National Academies.

Hammerle, M., Haynes, M., & McNeil, S. (2005). Use of Automatic Vehicle Location and Passenger Count Data to Evaluate Bus Operations. *Journal of the Transportation Research Record*, 27-34.

Hranac, R., Kwon, J., Bachmann, M., & Petty, K. (2011). Using Marey Graphs to Visualize Transit Loading and Schedule Adherence. *The 90th Annual Meeting of Transportation Research Board.* Washington, D.C.: Transportation Research Board of National Academies.

Kimpel. (2001). *Ph.D. Dissertation: Time Point-Level Analysis of Transit Service Reliability and Passenger Demand.* Portland, OR.

Kimpel, T. (2006). *Data Visualization as a Tool for Improved Decision Making within Transit Agencies.* Transportation Northwest.

Kimpel, T. J., Strathman, J. G., & Callas, S. (2008). Improving Scheduling Through Performance Monitoring. *Computer-Aided Systems in Public Transport*, 253-280.

Levinson, H. (1991). *Synthesis of Transit Practice 15: Supervision strategies for improved reliability of bus routes.* Washington D.C.: Transportation Research Record.

Liao, C. (2011). *Data-driven Support Tools for Transit Data Analysis, Scheduling and Planning.* Minneapolis: University of Minnesota at Twin Cities.

Liao, C., & Liu, H. (2010). Development of a Data Processing Framework for Transit Performance Analysis. *The 89th Annual Meeting of Transportation Research Board.* Washington, D.C.: Transportation Research Board of National Academies.

Mandelzys, M., & Hellinga, B. (2010). Automatically Identifying the Causes of Bus Transit Schedule Adherence Performance Issues Using AVL/APC Archived Data. *Transportation Research Board of the National Academies.* Washington D.C.

Strathman, J. G., Kimpel, T. J., Dueker, K. J., Gerhart, R. L., & Callas, S. (2002). Evaluation of Transit Operations: Data Applications of Tri-Met's Automated Bus Dispatching System. *Transportation*, 321-345.

Strathman, J., Dueker, K., Kimpel, T., Gerhart, R., Turner, K., & Taylor, P. (1999). Automated Bus Dispatching, Operations, Control, and Service Reliability. *Journal of Transportation Research Record*, 28-36.

Strathman, J., Dueker, K., Kimpel, T., Gerhart, R., Turner, K., & Taylor, P. (2000). Service Reliability Impacts of Computer-aided Dispatching and Automatic Vehicle Location Technology: A Tri-Met Case Study. *Transportation Quarterly*, 85-102.

Strathman, J., Dueker, K., Kimpel, T., Gerhart, R., Turner, K., Callas, S., et al. (2001). Review and an Experiment Involving Tri-Met's Automated Bus Dispatch System. *Journal of Public Transportation*, 1-26.

Strathman, J., Kimpel, T., & Callas, S. (2003). *Headway Deviation Effects on Bus Passenger Loads: Analysis of Tri-Met's Archived AVL-APC Data.* Portland, Oregon.

TCQSM 2nd Eds. (2003). *TCRP Report 100: Transit Capacity and Quality of Service Manual 2nd Edition.* Washington, D.C.: Transportation Research Record of the National Academies.

Turnquist, M. (1981). Strategies for Improving Reliability of Bus Transit Service. *Journal of Transportation Research Record*, 7-13.

P.O. Box 751
Portland, OR 97207

OTREC is dedicated to stimulating and conducting collaborative multi-disciplinary research on multi-modal surface transportation issues, educating a diverse array of current practitioners and future leaders in the transportation field, and encouraging implementation of relevant research results.